STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number 183000

TO: Bradley L Sisson
Location: rem/2B87/2C70
Art Unit: 1634
Friday, May 12, 2006
Case Serial Number: 09/423093

From: Toby Port
Location: Biotech-Chem Library
REM-1A59
Phone: (571)272-2523

toby.port@uspto.gov

Search Notes

Dear Examiner Sisson,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Toby Port
Technical Information Specialist
STIC Biotech/Chem Library
(571)272-2523

THIS PAGE BLANK (USPTO)

68940

188000

mg

STIC-Biotech/ChemLib

From: Sisson, Bradley
Sent: Monday, May 01, 2006 10:47 AM
To: STIC-Biotech/ChemLib
Subj ct: 09/423093

Please conduct a search of commercial, pending and allowed databases for :

oligos corresponding to positions 739-1932, 8646 to 9911, 9901-10953, and 11821-12945 of SEQ ID NO. 1

oligos corresponding to positions 79-861, 2011-2757, 5365-6471, 13156-13821, 2744-3109, and 858-2042 of SEQ ID NO: 2

oligos corresponding to positions 2352-3314, 3361-3875, 3977 to 5020, 6313-7323, 7310-8467, 1019-2359, 5114-6313 of SEQ ID NO: 3,

oligos corresponding to positions 12762-14054 and 14059-15060 of SEQ ID NO: 4.

Many thanks for your efforts!

Bradley L. Sisson
Primary Examiner
Art Unit 1634
REM-2B87
Mail Drop: REM 2C70
(V) 571-272-0751
(F) 571-273-0751

****
Searcher:_____
Searcher Phone:_____
Date Searcher Picked up:_____
Date completed:_____
Searcher Prep Time:_____
Online Time:_____

****
Type of Search
NA#_____ AA#:_____
S/L:_____ Oligomer:_____
Encode/Transl:_____
Structure #:_____Text:____
Inventor:_____ Litigation:___

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:_____
WWW/Internet:_____
Other (Specify):_____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:45:22 ; Search time 4333.97 Seconds
(without alignments)
15660.244 Million cell updates/sec

Title: US-09-423-093B-1_COPY_739_1932
Perfect score: 1194
Sequence: 1 atgttactatgctgcataca.....aatcaaataatgatttttga 1194

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1194	100.0	14516	1	AF078736	AF078736 Escherich
	2	1194	100.0	14516	6	BD085554	BD085554 Nucleic a
	3	1194	100.0	14520	6	AX029261	AX029261 Sequence
	4	875.4	73.3	13650	1	AF285969	AF285969 Salmonell
	5	305	25.5	1507	1	AY763519	AY763519 Shigella
	6	303.4	25.4	2259	1	S73325	S73325 Shigella dy
	7	301.8	25.3	1776	1	SHFRFPA	M96064 Shigella dy
	8	252.6	21.2	12033	1	AY532664	AY532664 Escherich
	9	223.8	18.7	340806	1	NMA1Z2491	AL162752 Neisseria
	10	223	18.7	1582	1	AF485419	AF485419 Neisseria
c	11	222.6	18.6	110000	1	AE002098_02	Continuation (3 of
c	12	222.6	18.6	349980	6	AX043922	AX043922 Sequence
	13	219.4	18.4	2963	1	NMU73942	U73942 Neisseria m
	14	217.8	18.2	1585	1	AF485418	AF485418 Neisseria
	15	217.8	18.2	110000	1	AE004969_17	Continuation (18 o
	16	170	14.2	1173	6	AR318240	AR318240 Sequence
c	17	165.6	13.9	110000	1	BA000039_12	Continuation (13 o
	18	132	11.1	110000	1	CR543861_00	CR543861 Acinetoba
c	19	123	10.3	110000	1	BA000028_29	Continuation (30 o
	20	122.2	10.2	14888	1	AE004568	AE004568 Pseudomon
	21	118	9.9	15155	1	AY208937	AY208937 Escherich
	22	116.4	9.7	17433	1	AY380835	AY380835 Shigella
	23	111.4	9.3	19457	1	AF461769	AF461769 Yersinia
	24	111.4	9.3	20125	1	AF461770	AF461770 Yersinia
	25	111.4	9.3	20511	1	AF461768	AF461768 Yersinia
	26	109.6	9.2	2081	1	PRMCPSFA	L36873 Proteus mir
	27	105	8.8	110000	1	CP000083_05	Continuation (6 of
c	28	102.2	8.6	110000	1	CR522870_00	CR522870 Desulfota
	29	91.8	7.7	110000	1	AE016828_07	Continuation (8 of
c	30	90	7.5	9138	1	AE015755	AE015755 Shewanell
	31	90	7.5	45993	1	AB012957	AB012957 Vibrio ch
	32	89.2	7.5	46392	1	U10927	U10927 Staphylococ
	33	89.2	7.5	46721	1	AB012956	AB012956 Vibrio ch
	34	88.4	7.4	1092	11	AY774474	AY774474 Synthetic
	35	88.4	7.4	1197	11	AY967147	AY967147 Synthetic
	36	88.4	7.4	17623	6	AX960586	AX960586 Sequence
	37	88.4	7.4	22401	1	AY101577	AY101577 Francisel
	38	88.4	7.4	22401	6	AX960626	AX960626 Sequence
c	39	88.4	7.4	110000	1	AJ749949_15	Continuation (16 o
c	40	85.2	7.1	110000	1	BA000004_38	Continuation (39 o
	41	81	6.8	110000	1	CP000082_07	Continuation (8 of
	42	78.4	6.6	18900	1	AY528413	AY528413 Escherich
	43	78.2	6.5	13171	1	VCU47057	U47057 Vibrio chol
	44	77	6.4	10029	1	AE013343	AE013343 Methanosa
	45	75.6	6.3	8594	1	AE013034	AE013034 Thermoana

ALIGNMENTS

RESULT 1
AF078736
LOCUS AF078736 14516 bp DNA linear BCT 03-AUG-1999
DEFINITION Escherichia coli O111 O antigen gene cluster, partial sequence.
ACCESSION AF078736 U13629
VERSION AF078736.1 GI:5678713
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 14516)
AUTHORS Bastin,D.A. and Reeves,P.R.
TITLE Sequence and analysis of the O antigen gene (rfb) cluster of Escherichia coli O111
JOURNAL Gene 164 (1), 17-23 (1995)
PUBMED 7590310
REFERENCE 2 (bases 3021 to 9981)
AUTHORS Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C., Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
TITLE Bacterial polysaccharide synthesis and gene nomenclature
JOURNAL Trends Microbiol. 4 (12), 495-503 (1996)
PUBMED 9004408
REFERENCE 3 (bases 1 to 14516)
AUTHORS Wang,L., Curd,H., Qu,W. and Reeves,P.R.
TITLE Sequencing of Escherichia coli O111 O-antigen gene cluster and identification of O111-specific genes
JOURNAL J. Clin. Microbiol. 36 (11), 3182-3187 (1998)
PUBMED 9774562
REFERENCE 4 (bases 3021 to 9981)
AUTHORS Bastin,D.A.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-1997) Microbiology, University of Sydney, Maze Crescent, Sydney, NSW 2006, Australia
REFERENCE 5 (bases 1 to 14516)
AUTHORS Wang,L., Curd,H., Qu,W., Bastin,D.A. and Reeves,P.R.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-1998) Microbiology, University of Sydney, Maze Crescent, Sydney, NSW 2006, Australia
COMMENT On Aug 3, 1999 this sequence version replaced gi:1155016.
FEATURES Location/Qualifiers

```
source          1. .14516
                /organism="Escherichia coli"
                /mol_type="genomic DNA"
                /strain="Stoke W; M92"
                /serotype="O111"
                /db_xref="taxon:562"
gene            <1. .273
                /gene="galF"
CDS             <1. .273
                /gene="galF"
                /codon_start=1
                /transl_table=11
                /product="putative glucose-1-phosphate
                uridylyltransferase"
                /protein_id="AAD46727.1"
                /db_xref="GI:5678714"
                /translation="DLMAVGRYVLSADIWAELEKTAPGAWGRIQLTDAIAELAKKQSV
                DAMLMTGDSYDCGKKMGYMQAFVKYGLRNLKEGAKFRKSIKKLLSE"
gene            739. .1932
                /gene="wbdH"
CDS             739. .1932
                /gene="wbdH"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase"
                /protein_id="AAD46728.1"
                /db_xref="GI:5678715"
                /translation="MLLCCIHINVYYLLLECDMKKIVIIGNVASMMLRFRKELIMNLV
                RQGDNVYCLANDFSTEDLKVLSSWGVKGVKFSLNSKGINPFKDIIAVYELKKILKDIS
                PDIVFSYFVKPVIFGTIASKLSKVPRIVGMIEGLGNAFTYYKGKQTTKTKMIKWIQIL
                LYKLALPMLDDLILLNHDDKKDLIDQYNIKAKVTVLGGIGLDLNEFSYKEPPKEKITF
                IFIARLLREKGIFEFIEAAKFVKTTYPSSEFVILGGFESNNPFSLQKNEIESLRKEHD
                LIYPGHVENVQDWLEKSSVFVLPTSYREGVPRVIQEAMAIGRPVITTNVPGCRDIIND
                GVNGFLIPPFEINLLAEKMKYFIENKDKVLEMGLAGRKFAEKNFDAFEKNNRLASIIK
                SNNDF"
gene            1983. .3101
                /gene="gmd"
CDS             1983. .3101
                /gene="gmd"
                /codon_start=1
                /transl_table=11
                /product="GDP-D-mannose dehydratase"
                /protein_id="AAD46729.1"
                /db_xref="GI:5678716"
                /translation="MNKVALITGITGQDGSYLAELLLEKGYEVHGIKRRASSFNTERV
                DHIYQDSHLANPKLFLHYGDLTDTSNLTRILKEVQPDEVYNLGAMSHVAVSFESPEYT
                ADVDAIGTLRLLEAIRILGLEKKTKFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
                KLYAYWITVNYRESYGMFACNGILFNHESPRRGETFVTRKITRGIANIAQGLDKCLYL
                GNMDSLRDWGHAKDYVKMQWMMLQQETPEDFVIATGIQYSVREFVTMAAEQVGIELAF
                EGEGVNEKGVVVSVNGTDAKAVNPGDVIISVDPRYFRPAEVETLLGDPTNAHKKLGWS
                PEITLREMVKEMVSSDLAIAKKNVLLKANNIATNIPQE"
gene            3174. .3623
                /gene="wbdI"
CDS             3174. .3623
                /gene="wbdI"
                /note="previously known as orf3.4"
                /codon_start=1
                /transl_table=11
                /product="WbdI"
                /protein_id="AAC44881.1"
                /db_xref="GI:1155017"
                /translation="MFITSDKFREIIKLVPLVSIDLLIENENGEYLFGLRNNRPAKNY
                FFVPGGRIRKNESIKNAFKRISSMELGKEYGISGSVFNGVWEHFYDDGFFSEGEATHY
                IVLCYTLKVLKSELNLPDDQHREYLWLTKHQINAKQDVHNYSKNYFL"
gene            3657. .5108
                /gene="manC"
CDS             3657. .5108
                /gene="manC"
                /EC_number="2.7.7.22"
                /note="previously known as RfbM"
                /codon_start=1
                /transl_table=11
                /product="mannose-1-phosphate guanylyltransferase"
                /protein_id="AAC44882.1"
                /db_xref="GI:1155018"
                /translation="MSQCLYPVIIAGGTGSRLWPLSRVLYPKQFLNLVGDSTMLQTTI
                TRLDGIECENPIVICNEDHRFIVAEQLRQIGKLTKNIILEPKGRNTAPAIALAAFIAQ
                KNNPNDDPLLLVLAADHSINNEKAFRESIIKAMPYATSGKLVTFGIIPDTANTGYGYI
                KRSSSADPNKEFPAYNVAEFVEKPDVKTAQEYISSGNYYWNSGMFLFRASKYLDELRK
                FRPDIYHSCECATATANIDMDFVRINEAEFINCPEESIDYAVMEKTKDAVVLPIDIGW
                NDVGSWSSLWDISQKDCHGNVCHGDVLNHDGENSFIYSESSLVATVGVSNLVIVQTKD
                AVLVADRDKVQNVKNIVDDLKKRKRAEYYMHRAVFRPWGKFDAIDQGDRYRVKKIIVK
                PGEGLDLRMHHHRAEHWIVVSGTAKVSLGSEVKLLVSNESIYIPQGAKYSLENPGVIP
                LHLIEVSSGDYLESDDIVRFTDRYNSKQFLKRD"
gene            5114. .6484
                /gene="manB"
CDS             5114. .6484
                /gene="manB"
                /EC_number="5.4.2.8"
                /note="previously known as RfbK"
                /codon_start=1
                /transl_table=11
                /product="phosphomannomutase"
                /protein_id="AAC44883.1"
                /db_xref="GI:1155019"
                /translation="MNKITCFKAYDIRGRLGAELNDEIAYRIGRAYGEFFKPQTVVVG
                GDARLTSESLKKSLSNGLCDAGVNVLDLGMCGTEEIYFSTWYLGIDGGIEVTASHNPI
                DYNGMKLVTKGARPISSDTGLKDIQQLVESNNFEELNLEKKGNITKYSTRDAYINHLM
                GYANLQKIKKIKIVVNSGNGAAGPVIDAIEECFLRNNIPIQFVKINNTPDGNFPHGIP
                NPLLPECREDTSSAVIRHSADFGIAFDGDFDRCFFFDENGQFIEGYYIVGLLAEVFLG
                KYPNAKIIHDPRLIWNTIDIVESHGGIPIMTKTGHAYIKQRMREEDAVYGGEMSAHHY
                FKDFAYCDSGMIPWILICELLSLTNKKLGELVCGCINDWPASGEINCTLDNPQNEIDK
                LFNRYKDSALAVDYTDGLTMEFSDWRFNVRCSNTEPVVRLNVESRNNAILMQEKTEEI
                LNFISK"
gene            6521. .7444
                /gene="wbdJ"
CDS             6521. .7444
                /gene="wbdJ"
                /note="previously known as orf6.7"
                /codon_start=1
                /transl_table=11
                /product="WbdJ"
                /protein_id="AAC44884.1"
                /db_xref="GI:1155020"
                /translation="MKVLLTGSTGMVGKNILEHDSASKYNILTPTSSDLNLLDKNEIE
                KFMLINMPDCIIHAAGLVGGIHANISRPFDFLEKNLQMGLNLVSVAKKLGIKKVLNLG
                SSCMYPKNFEEAIPEKALLTGELEETNEGYAIAKIAVAKACEYISRENSNYFYKTIIP
                CNLYGKYDKFDDNSSHMIPAVIKKIHHAKINNVPEIEIWGDGNSRREFMYAEDLADLI
                FYVIPKIEFMPNMVNAGLGYDYSINDYYKIIAEEIGYTGSFSHDLTKPTGMKRKLVDI
                SLLNKIGWSSHFELRDGIRKTYNYYLENQNK"
gene            7441. .8607
                /gene="wbdK"
CDS             7441. .8607
                /gene="wbdK"
                /note="previously known as orf7.7"
                /codon_start=1
                /transl_table=11
                /product="WbdK"
                /protein_id="AAC44885.1"
                /db_xref="GI:1155021"
                /translation="MITYPLASNTWDEYEYAAIQSVIDSKMFTMGKKVELYEKNFADL
                FGSKYAVMVSSGSTANLLMIAALFFTNKPKLKRGDEIIVPAVSWSTTYYPLQQYGLKV
                KFVDINKETLNIDIDSLKNAISDKTKAILTVNLLGNPNDFAKINEIINNRDIILLEDN
                CESMGAVFQNKQAGTFGVMGTFSSFYSHHIATMEGGCVVTDDEELYHVLLCLRAHGWT
                RNLPKENMVTGTKSDDIFEESFKFVLPGYNVRPLEMSGAIGIEQLKKLPGFISTRRSN
                AQYFVDKFKDHPFLDIQKEVGESSWFGFSFVIKEGAAIERKSLVNNLISAGIECRPIV
                TGNFLKNERVLSYFDYSVHDTVANAEYIDKNGFFVGNHQIPLFNEIDYLRKVLK"
gene            8646. .9911
                /gene="wzx"
CDS             8646. .9911
                /gene="wzx"
                /note="integral membrane protein; previously known as
                orf8.9"
                /codon_start=1
                /transl_table=11
                /product="putative O unit flippase"
```

```
                     /protein_id="AAC44886.1"
                     /db_xref="GI:1155022"
                     /translation="MVLTVKKILAFGYSKVLPPVIEQFVNPICIFIITPLILNHLGKQ
                     SYGNWILLITIVSFSQLICGGCSAWIAKIIAEQRILSDLSKKNALRQISYNFSIVIIA
                     FAVLISFLILSICFFDVARNNSSFLFAIIICGFFQEVDNLFSGALKGFEKFNVSCFFE
                     VITRVLWASIVIYGIYGNALLYFTCLAFTIKGMLKYILVCLNITGCFINPNFNRVGIV
                     NLLNESKWMFLQLTGGVSLSLFDRLVIPLILSVSKLASYVPCLQLAQLMFTLSASANQ
                     ILLPMFARMKASNTFPSNCFFKILLVSLISVLPCLALFFFGRDILSIWINPTFATENY
                     KLMQILAISYILLSMMTSFHFLLLGIGKSKLVANLNLVAGLALAASTLIAAHYGLYAI
                     SMVKIIYPAFQFYYLYVAFVYFNRAKNVY"
     gene            9901. .10953
                     /gene="wzy"
     CDS             9901. .10953
                     /gene="wzy"
                     /codon_start=1

  Query Match             100.0%;  Score 1194;  DB 1;  Length 14516;
  Best Local Similarity   100.0%;  Pred. No. 8e-171;
  Matches 1194;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 798

Qy         61 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 858

Qy        121 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        859 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 918

Qy        181 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        919 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAG 978

Qy        241 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        979 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 1038

Qy        301 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1039 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 1098

Qy        361 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1099 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 1158

Qy        421 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1159 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 1218

Qy        481 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1219 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 1278

Qy        541 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1279 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 1338

Qy        601 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1339 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 1398

Qy        661 ATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1399 ATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTT 1458

Qy        721 AAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1459 AAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCT 1518

Qy        781 TTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1519 TTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCT 1578

Qy        841 GGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1579 GGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACA 1638

Qy        901 TCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1639 TCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTA 1698

Qy        961 ATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1699 ATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTG 1758

Qy       1021 ATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1759 ATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAA 1818

Qy       1081 GATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1819 GATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCT 1878

Qy       1141 TTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1879 TTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1932


RESULT 2
BD085554
LOCUS       BD085554               14516 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085554
VERSION     BD085554.1  GI:22631164
KEYWORDS    JP 2001522248-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 14516)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 1 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Escherichia coli
            PN   JP 2001522248-A/1
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14516
            FT                   /organism='Escherichia coli'.
FEATURES             Location/Qualifiers
     source          1. .14516
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 1194;  DB 6;  Length 14516;
  Best Local Similarity   100.0%;  Pred. No. 8e-171;
  Matches 1194;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 798

Qy         61 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        799 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 858
Qy        121 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 180
Db        859 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 918
Qy        181 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAG 240
Db        919 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAG 978
Qy        241 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 300
Db        979 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 1038
Qy        301 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 360
Db       1039 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 1098
Qy        361 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 420
Db       1099 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 1158
Qy        421 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 480
Db       1159 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 1218
Qy        481 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 540
Db       1219 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 1278
Qy        541 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 600
Db       1279 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 1338
Qy        601 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 660
Db       1339 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 1398
Qy        661 ATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTT 720
Db       1399 ATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTT 1458
Qy        721 AAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCT 780
Db       1459 AAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCT 1518
Qy        781 TTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCT 840
Db       1519 TTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCT 1578
Qy        841 GGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACA 900
Db       1579 GGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACA 1638
Qy        901 TCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTA 960
Db       1639 TCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTA 1698
Qy        961 ATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTG 1020
Db       1699 ATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTG 1758
Qy       1021 ATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAA 1080
Db       1759 ATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAA 1818
Qy       1081 GATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCT 1140
Db       1819 GATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCT 1878
Qy       1141 TTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1194
Db       1879 TTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1932
```

```
RESULT 3
AX029261
LOCUS       AX029261               14520 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 1 from Patent WO9850531.
ACCESSION   AX029261
VERSION     AX029261.1  GI:10190112
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 1 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .14520
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
ORIGIN

  Query Match          100.0%;  Score 1194;  DB 6;  Length 14520;
  Best Local Similarity  100.0%;  Pred. No. 8e-171;
  Matches 1194;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 60
Db        739 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 798
Qy         61 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 120
Db        799 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 858
Qy        121 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 180
Db        859 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 918
Qy        181 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAG 240
Db        919 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAG 978
Qy        241 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 300
Db        979 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 1038
Qy        301 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 360
Db       1039 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 1098
Qy        361 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 420
Db       1099 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 1158
Qy        421 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 480
Db       1159 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 1218
Qy        481 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 540
Db       1219 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 1278
Qy        541 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 600
Db       1279 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 1338
Qy        601 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 660
Db       1339 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 1398
```

```
Qy        661 ATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1399 ATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTT 1458

Qy        721 AAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1459 AAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCT 1518

Qy        781 TTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1519 TTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCT 1578

Qy        841 GGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1579 GGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACA 1638

Qy        901 TCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1639 TCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTA 1698

Qy        961 ATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1699 ATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTG 1758

Qy       1021 ATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1759 ATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAA 1818

Qy       1081 GATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1819 GATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCT 1878

Qy       1141 TTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1879 TTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1932


RESULT 4
AF285969
LOCUS       AF285969               13650 bp    DNA     linear   BCT 02-AUG-2001
DEFINITION  Salmonella enterica O antigen gene cluster, complete sequence.
ACCESSION   AF285969
VERSION     AF285969.1  GI:15076983
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 13650)
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     The E. coli O111 and S. enterica O35 gene clusters: Gene clusters
            encoding the same colitose containing O antigen are highly
            conserved
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 13650)
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2000) Department of Microbiology, The University
            of Sydney, Sydney, NSW 2006, Australia
FEATURES             Location/Qualifiers
     source          1. .13650
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /serovar="Adelaide"
                     /db_xref="taxon:28901"
                     /note="serogroup: O35"
     gene            1321. .2460
                     /gene="wbdH"
     CDS             1321. .2460
                     /gene="wbdH"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glycosyl transferase WbdH"
                     /protein_id="AAK83009.1"
                     /db_xref="GI:15076984"
                     /translation="MKKIVIIGNIASMMLRFRKELIINLVSQGDDVYCLANDFSPEDL
                     KLLSSWGGKGIKYTLNSKGINPFKDILSVIELKNILDKISPDIVFSYFVKPVIFGTIA
                     SKLSRVPRIVGMIEGLGNAFTYYKGKQSIKTRIIKWMQIFLYKLALPMLNDLVLLNKD
                     DKKDLIERYNIKVNVTILGGIGLNLNDFSYKEPLTDKVTFIFIARLLREKGVFEYIEA
                     AKRVKNIYPNVEFVILGGFENNNPFSLQQHEIELLRKEHDLIYPGHVENVQDWLEKSS
                     VFVLPTSYREGVPRVIQEAMAIGRPVITTNVPGCKDIINDGINGFLVPPFESELLVKK
                     MIYFIDNRSKILEMGLSGRLFAEKNFDAMEKNKTLASIIKANHEF"
     gene            2511. .3629
                     /gene="gmd"
     CDS             2511. .3629
                     /gene="gmd"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-D-mannose dehydratase Gmd"
                     /protein_id="AAK83010.1"
                     /db_xref="GI:15076985"
                     /translation="MNKVALITGITGQDGSYLAELLLEKGYEVHGIKRRASSFNTERV
                     DHIYQDIHLNNPKLYLHYGDLTDTSNLTRILQEVQPDEVYNLGAMSHVAVSFESPEYT
                     ADVDALGTLRLLEAIRILGLEKKTKFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
                     KLYAYWITVNYRESYGMFACNGILFNHESPRRGETFVTRKITRGIANIAQGLDKCLYL
                     GNMDSLRDWGHAKDYVKMQWMMLQQEAPEDFVIATGVQYSVRDFVTMAAKQVGIELVF
                     EGEGINEKGIVASVNGNDAKAVKAGDVIISVDPRYFRPAEVETLLGDPTNAHKKLGWS
                     PEITLCEMVKEMVSSDLAIAKKNVLLKTNNIPTNIPQE"
     gene            3698. .4147
                     /gene="gmm"
     CDS             3698. .4147
                     /gene="gmm"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-mannosyl hydrolase Gmm"
                     /protein_id="AAK83011.1"
                     /db_xref="GI:15076986"
                     /translation="MFIPSDKFREVIKLTPLVSIDILIENENGEYLFGLRNNRPAKNQ
                     FFVPGGRIRKSEFIENAFKRISFAELGKEYDISESLFNGIWEHFYEESFFSDADATHY
                     IVLCYKLKVLKNELNLPADQHCEYIWISEDKISNLNNIHKYSKDYFL"
     gene            4181. .5626
                     /gene="manC"
     CDS             4181. .5626
                     /gene="manC"
                     /codon_start=1
                     /transl_table=11
                     /product="mannose-1-phosphate guanylyltransferase ManC"
                     /protein_id="AAK83012.1"
                     /db_xref="GI:15076987"
                     /translation="MSQCLYPVVIAGGTGSRLWPLSRVLYPKQFLNLVGNSTMLQTTI
                     ARLNGIKCENPIVICNEDHRFIVAEQLRQIDKLTKNIILEPKGRNTAPAIALAAFIAQ
                     KKKPKDDPLLLVLAADHSINNEQAFREAIIKAIPYADAGKLVMFGIIPSVAHTGYGYI
                     KRSVPVDADKDSTAYYVANFVEKPNIQKAQEYIMSEGYYWNSGMFLFRASKYLGELKK
                     FRPDIYNKCESATATANIDMDFVRINETEFINCPEESIDYAVMEKTKDAVVLPIDVGW
                     SDVGSWSSLWDISQKDVHGNVCQGDVINHDSEDSFIYSESSLVATVGVSNLVIVQTKD
                     AVLVADRDKVQNVKNIVDSLKQGGRAEYYMHREVFRPWGKYDAIDQGDRYKVKKIIVK
                     PGEGLDLRMHHHRAEHWIVVSGTAKVSLGNKIKLLVPNESIYIPQGAEYSLENPGVIP
                     LHLIEVSSGDYLESDDIVRFTDRYNSKKFLK"
     gene            5639. .7015
                     /gene="manB"
     CDS             5639. .7015
                     /gene="manB"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphomannomutase ManB"
                     /protein_id="AAK83013.1"
                     /db_xref="GI:15076988"
                     /translation="MNKLTCFKAYDIRGRLGVELDDDIAYRIGRAYGELFKPQTVVVG
                     GDTRLTSESLKKSLSNGLCDAGVNVLDLGMCGTEEIYFSTWHLGVDGGIEVTASHNPI
                     DYNGMKLVTKNARPISSDTGLKDIQTLVELNDFEKINLEKTGNITKYSTRDAYIKHLM
                     GYVNLQKIKKIKIVVNSGNGAAGPVIDDIEKYFLQNNIPVTFVKINNTPDGHFPNGIP
                     NPLLPECREDTSKAVIAHSADFGIAFDGDFDRCFFFDEKGRFIEGYYIVGLLAEVFLE
                     KYPNAKIIHDPRLIWNTIDIVQNYGGKPIMTKTGHAYIKQRMREEDAVYGGEMSAHHY
                     FKDFEYCDSGMIPWILICELLSIRNKNLGELVYNCINAWPASGEINCTLSNPQSEIDK
                     LFFYYKNSAISVDYTDGLSMEFSDWRFNVRCSNTEPVVRLNVESKSNVLLMQEKTREI
```

```
                LSFIGKCV"
gene            7040. .7963
                /gene="wbdJ"
CDS             7040. .7963
                /gene="wbdJ"
                /codon_start=1
                /transl_table=11
                /product="WbdJ"
                /protein_id="AAK83014.1"
                /db_xref="GI:15076989"
                /translation="MKVLLTGSSGMVGKNILEHASAKDYEILTPTSSDLNLLVIDEIE
                KFMLANKPECIIHAAGLVGGIHANINRPFDFLANNLLMGLNLVTIAKKLGIKKFLNLG
                SSCMYPRNFEEAIPEKALLTGELEETNEGYAIAKVAVAKACEYISRENSNYYYKTIIP
                CNLYGKYDKFGDDSSHMIPAVIKKIHHAKANNIPEIEIWGDGNSRREFMYAEDLADLV
                FYIIPKIESIPNMVNAGLGYDYTVNDYYKIIAEEIGYTGTFFHDLTKPTGMKRKLVDI
                SLLNKIGWSSHPNLRDGIRKTYNYYLENQYK"
gene            7960. .9126
CDS             7960. .9126
                /gene="wbdK"
                /codon_start=1
                /transl_table=11
                /product="WbdK"
                /protein_id="AAK83015.1"
                /db_xref="GI:15076990"
                /translation="MITYPLASNTWDENEYAAIQSVIDSKMFTMGKNVELYEKNFADL
                FGSKYAVMVSSGSTANLLMIAALFFTHKPKLKRGDEIIVPAVSWSTTYYPLQQYGLKV
                KFVDINKETLNIDVDSLKNAISDKTRAILAVNLLGNPNDFSTIHEIISNKDIILLEDN
                CESMGATFQNKQAGTFGIMGSFSSFYSHHIATMEGGCVVTDDEELYHILLCLRAHGWT
                RNLPKENMVTGTKSDDIFEESFKFVLPGYNVRPLEMSGAIGIEQLKKLPGFISTRRSN
                AQYFVDKFKDHPFLDIQKEIGESSWFGFSLIIKQDTGIERKSLINKLITSGIECRPIV
                TGNFLKNKRVLSYFDYSVHDTVVNAEYIDKNGFFVGNHQIPLLNEIDFLRKVLK"
gene            9178. .10431
                /gene="wzx"
CDS             9178. .10431
                /gene="wzx"
                /codon_start=1
                /transl_table=11
                /product="O antigen flippase Wzx"
                /protein_id="AAK83016.1"
                /db_xref="GI:15076991"
                /translation="MKKFLAFGYSKVLPPIIEQFINPICIFVITPLILNHLGKQSYGN
                WILLITIVSFSQLICGGCSAWIAKVIAEQRATSDYAGKGALRKIAYNLSIIMLSFAIL
                ILLILLSLCVFDVARDNTEFLYAIAICGFFQEVDNLFNGALKGFERFNVSCLFEVITR
                VLWAFIVIYGIQENSVLYYTCLAFILKGTLKYILVCLNISGKFIYPNFDKAGIVNLLN
                ESKWMFLQLTGGISLGLFDRLVIPLILSVNKLSSYVPCLQLVQLMFTISASANQILLP
                MFARMKASNTFPSNCFFKIQLVSLSSALPCLVLLFFGYDILSIWISPAFASENFKLMQ
                ILSISYILLSMMTSFHFLLLGLGKSKLVANLNLIAGLTLAASTLIAAHYGLYAISVVK
                IIYPLFQFYYLYVAFFYFKREKNVY"
gene            10421. .11485
CDS             10421. .11485
                /gene="wzy"
                /codon_start=1
                /transl_table=11
                /product="O antigen polymerase Wzy"
                /protein_id="AAK83017.1"
                /db_xref="GI:15076992"
                /translation="MSIDFLFSITEIAVVFSCAIYIFVQCLLMHKIYLDKVVLILLCL
                LLCLVIIQLPELNEQGLIDAVKLLLPLLIVFIAFQKPKLCLWIIIALLFFNSTLNFWH
                FKVFDKFSSFPFTFFILLYFLYKLVNDQPPRQRNKKFYVVIFIFILIDVTQSLLISYR
                GQILYSLICVIMLVLRINLRKTIPYIFLTLPFIYVFIMTFIGFNYFNKDVVFFEPTAS
                NIERTGMIYYLVIHANDYIFHGMGTLHFLNEGGQYKMLYGFPSLIPNDPHDFLLRFFI
                SIGMMGALIYHCIFFVFFKNISSLLRQKNSYFVIVSCLLLLQVVLFYTLNPFNAFNRL
                ICGLTIGVIYGFSKSRQIYM"
gene            11451. .12344
                /gene="wbdL"
CDS             11451. .12344
                /gene="wbdL"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase WbdL"
                /protein_id="AAK83018.1"
                /db_xref="GI:15076993"
                /translation="MDFQRVDRYICSESFNTPLISIIIATFNSELVIANCLQSVISQS
                YKNIEVIIMDGGSYDKTLDIARSFKDERIKIVSEKDRGIYDAWNKAVDLSSGDWIAFI
                GSDDVYYHTDAITSLVKGAAISNGAPVVYGRTAHEGPNKEISGFSGSEWYNLKGFKFN
                YYKCNLPLPIMSAIYSRDFFKDQRFDIKLKIVADADWFLRCFIKWRKENSPYFIDDIT
                PVVRMGYGGVSTDISSQIKTTRESFIVRKKNNISCFNMQLILRYIKIMVMVSIKNIFG
                NNIYNSMHNGYHFLKKIKNKI"
gene            12341. .13465
                /gene="wbdM"
CDS             12341. .13465
                /gene="wbdM"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase WbdM"
                /protein_id="AAK83019.1"
                /db_xref="GI:15076994"
                /translation="MKILYVITGLTCGGAEHLMMQLADQMLLRGHDVNIICLSGISEI
                QPTQKHNIHYIRMEKNVRGFVKALFQVRKIITVIKPDIIHSHMFHANIFSRIIRILTP
                SVPLICTAHNKNEGGYARMLCYRLSDFLASITTNVSQEAVQEFIAKKAAPKNKIIEVP
                NFVNTTNFSFDLKARKERRNFFGIKDNTLLLLAVGRLVDAKDYPNLLNAINYLTLSES
                TGFDNFILFIAGDGVLKSKLVDLVNKLNLADKVFFLGRRSDIKELMCAADLFVLSSQW

Query Match             73.3%;  Score 875.4;  DB 1;  Length 13650;
Best Local Similarity   84.5%;  Pred. No. 7.5e-123;
Matches  995;  Conservative    0;  Mismatches  181;  Indels    1;  Gaps    1;

Qy         18 ACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAAAAAATAGTGATCATAGG 77
              | ||||   || |||||||||| |||||||||   ||||||||||||||||||||| |||||
Db       1285 ATATATTGTTGAATATTATTTA-TTTTAGAATTCAATATGAAAAAAATAGTGATTATAGG 1343

Qy         78 CAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATCATGAATTTAGTGAGGCA 137
               ||| |||| ||||||||| |||| || || |||||||||||||| ||| ||||||| ||
Db       1344 TAATATAGCATCAATGATGCTAAGATTTAGAAAAGAATTAATCATAAATCTAGTGAGCCA 1403

Qy        138 AGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAAGATCTTAAAGTACTTTC 197
              ||||||| |||| |||||||| |||||||||||||   | |||||| | ||| | |||||
Db       1404 AGGTGATGATGTTTATTGTCTTGCAAATGATTTTTCACCAGAAGATTTAAAATTGCTTTC 1463

Qy        198 GTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTTAA 257
              |||||||| | |||||| |||||| ||||||||||||||||||| ||||||||||
Db       1464 TTCATGGGGAGGTAAGGGTATTAAATATACTCTTAACTCAAAGGGTATAAATCCTTTTAA 1523

Qy        258 GGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGATATTTCCCCAGATATTGT 317
              ||||||| || | ||| |||||| |||| || ||| |  | || || |||||||| ||
Db       1524 GGATATACTTTCCGTTATTGAACTTAAAAATATCCTTGACAAAATATCTCCAGATATAGT 1583

Qy        318 ATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCTTCAAAGTTGTCAAAAGT 377
              |||||||||||||| |||||||||||| || || ||||||||||||| |||| |||
Db       1584 TTTTTCATATTTTGTGAAGCCAGTAATATTCGGGACCATTGCTTCAAAGTTATCAAGAGT 1643

Qy        378 GCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTCACTTATTATAAGGGAAA 437
              |||||||||||||||||| |||||||||| ||||| |||||||| ||||| |||||
Db       1644 ACCAAGGATTGTTGGAATGATCGAAGGTCTAGGAAATGCTTTCACTTACTATAAAGGAAA 1703

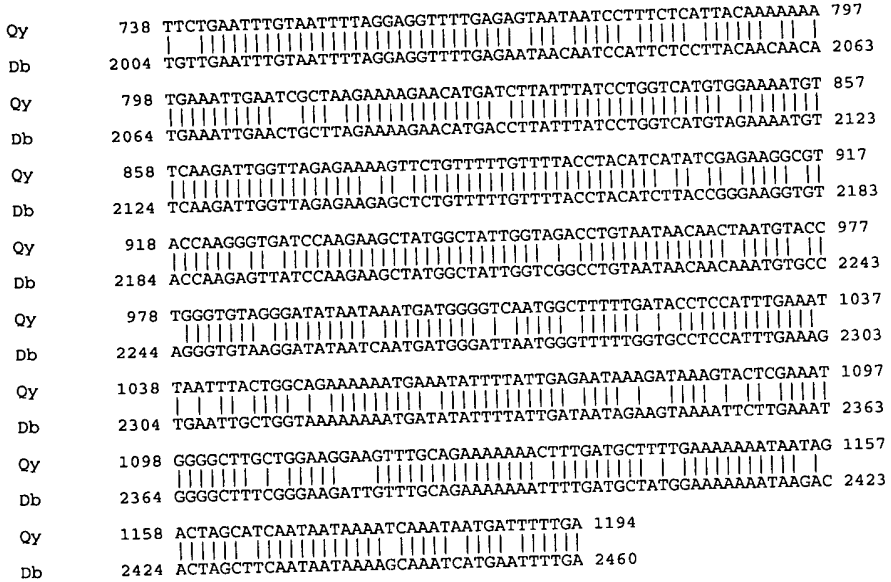
Qy        438 GCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATT 497
              |||| | | |||||||| ||| ||||| ||||| |||||| |||||||||||||| |||||
Db       1704 GCAGTCTATAAAAACTAGAATAATAAAATGGATGCAAATTTTTTTATATAAGTTGGCATT 1763

Qy        498 ACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCA 557
              ||||||||| |||||||| |||| |||||| | |||||||||||||| ||||||| |
Db       1764 GCCGATGCTTAATGATTTGGTTCTCTTAAATAAAGATGATAAAAAAGATCTAATCGAACG 1823

Qy        558 GTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTT 617
              |||||||||||| | ||| || ||| | |||| |||||||||| |||||||||| ||
Db       1824 GTATAATATTAAGGTTAATGTTACAATATTAGGCGGGATTGGATTAAATCTTAATGATTT 1883

Qy        618 TTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAG 677
              |||||||||||| |||| | ||| ||| |||| |||||||||||||| || || ||||
Db       1884 TTCATATAAAGAACCACTTACAGATAAAGTTACATTTATTTTTATAGCGAGACTACTAAG 1943

Qy        678 AGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAG 737
              |||||||||| ||||||| | ||| |||||||| ||     ||||||| | ||| ||| |
Db       1944 AGAGAAAGGGGTATTTGAATATATCGAAGCCGCGAAACGTGTTAAGAATATTTACCCGAA 2003
```

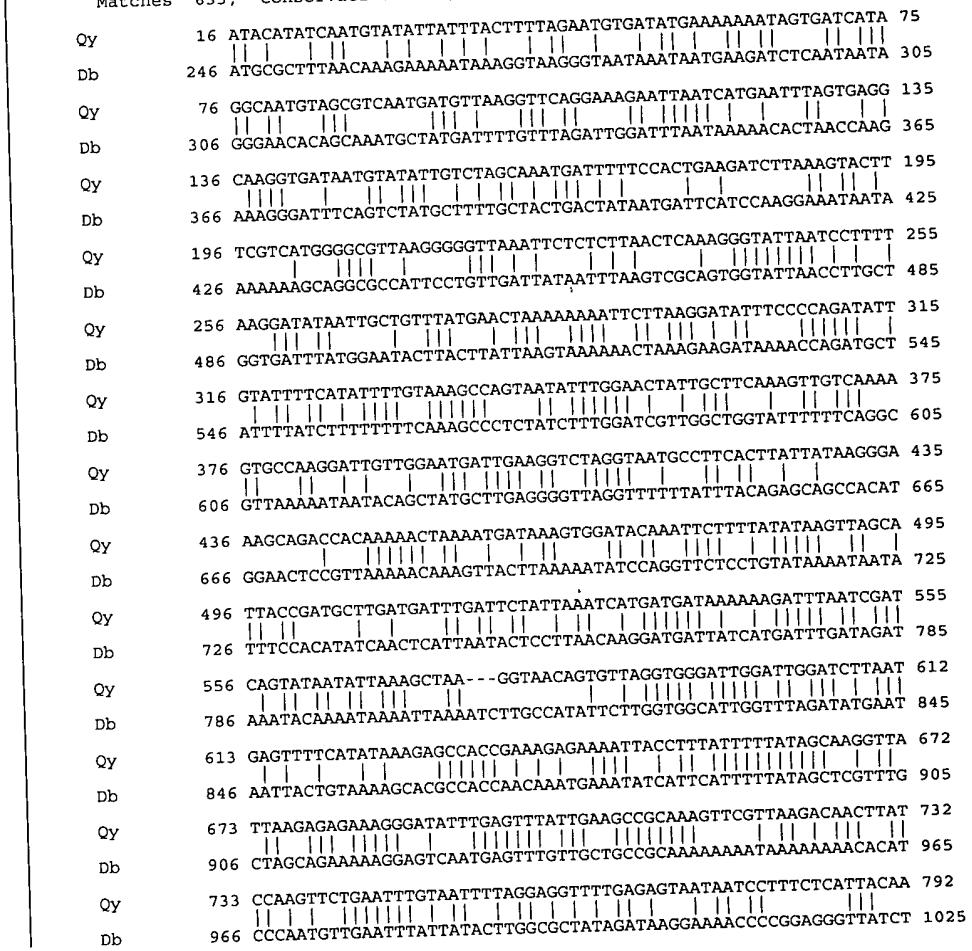


RESULT 5
AY763519
LOCUS AY763519 1507 bp DNA linear BCT 27-OCT-2004
DEFINITION Shigella dysenteriae glycosyl transferase (rfp) gene, complete cds.
ACCESSION AY763519
VERSION AY763519.1 GI:54402398
KEYWORDS .
SOURCE Shigella dysenteriae
ORGANISM Shigella dysenteriae
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.
REFERENCE 1 (bases 1 to 1507)
AUTHORS Xu,D.Q., Cisar,J.O. and Kopecko,D.J.
TITLE Molecular cloning, characterization and expression of the polysaccharide biosynthetic genes for Shigella dysenteriae type 1 in a live Salmonella vaccine vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1507)
AUTHORS Xu,D.Q., Cisar,J.O. and Kopecko,D.J.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-2004) Division of Bacterial, Parasitic and Allegenic Products, Laboratory of Enteric and Sexually Transmitted Diseases, Center for Biologics Evaluation and Research, Food and Drug Administration, 8800 Rockville Pike, Bethesda, MD 20892, USA

FEATURES Location/Qualifiers
source 1. .1507
/organism="Shigella dysenteriae"
/mol_type="genomic DNA"
/strain="1617"
/serotype="1"
/db_xref="taxon:622"
gene 226. .1421
/gene="rfp"
-35_signal 226. .231
/gene="rfp"
TATA_signal 235. .240
/gene="rfp"
RBS 275. .277
/gene="rfp"
CDS 288. .1421
/gene="rfp"
/note="similar to Shigella dysenteriae type 1 (strain 1) Rfp"
/codon_start=1
/transl_table=11
/product="glycosyl transferase"
/protein_id="AAV34754.1"
/db_xref="GI:54402399"
/translation="MKISIIGNTANAMILFRLDLIKTLTKKGISVYAFATDYNDSSKE
IIKKAGAIPVDYNLSRSGINLAGDLWNTYLLSKKLKKIKPDAILSFFSKPSIFGSLAG
IFSGVKNNTAMLEGLGFLFTEQPHGTPLKTKLLKNIQVLLYKIIFPHINSLILLNKDD
YHDLIDKYKIKLKSCHILGGIGLDMNNYCKSTPPTNEISFIFIARLLAEKGVNEFVAA
AKKIKKTHPNVEFIILGAIDKENPGGLSESDVDTLIKSGVISYPGFVSNVADWIEKSS
VFVLPSYYREGVPRSTQEAMAMGRPILTTNLPGCKETIIDGVNGYVVKKWSHEDLAEK
MLKLINNPEKIISMGEESYKLARERFDANVNNVKLLKILGIPD"

ORIGIN

Query Match 25.5%; Score 305; DB 1; Length 1507;
Best Local Similarity 54.5%; Pred. No. 1.3e-36;
Matches 633; Conservative 0; Mismatches 525; Indels 3; Gaps 1;



```
Qy        793 AAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAA 852
Db       1026 GAATCTGACGTAGATACTTTAATTAAATCAGGAGTTATTTCTTATCCCGGATTTGTTTCT 1085

Qy        853 AATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAA 912
Db       1086 AATGTGGCTGATTGGATTGAAAAATCAAGCGTATTTGTTCTTCCTTCCTATTATCGAGAG 1145

Qy        913 GGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAAT 972
Db       1146 GGAGTTCCTCGTAGTACACAAGAAGCGATGGCTATGGGGAGGCCGATTTTAACTACTAAT 1205

Qy        973 GTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTT 1032
Db       1206 TTACCAGGCTGCAAAGAAACAATTATTGATGGTGTGAATGGATATGTTGTAAAAAAATGG 1265

Qy       1033 GAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAAGTACTC 1092
Db       1266 TCACATGAAGATCTTGCAGAAAAAATGCTGAAGTTAATTAATAATCCTGAAAAAATAATC 1325

Qy       1093 GAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCTTTTGAAAAAAAT 1152
Db       1326 AGTATGGGAGAAGAAAGTTATAAGTTAGCAAGAGAAAGATTCGATGCAAATGTAAATAAT 1385

Qy       1153 AATAGACTAGCATCAATAATA 1173
Db       1386 GTAAAGTTATTAAAAATACTA 1406
```

RESULT 6
S73325
LOCUS S73325 2259 bp DNA linear BCT 30-SEP-2003
DEFINITION Shigella dysenteriae RfpA (rfpA) and RfpB (rfpB) genes, complete cds; and unknown gene.
ACCESSION S73325
VERSION S73325.1 GI:688320
KEYWORDS .
SOURCE Shigella dysenteriae
ORGANISM Shigella dysenteriae
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.
REFERENCE 1 (bases 1 to 2259)
AUTHORS Gohmann,S., Manning,P.A., Alpert,C.A., Walker,M.J. and Timmis,K.N.
TITLE Lipopolysaccharide O-antigen biosynthesis in Shigella dysenteriae serotype 1: analysis of the plasmid-carried rfp determinant
JOURNAL Microb. Pathog. 16 (1), 53-64 (1994)
PUBMED 7520113
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 155395] from the original journal article.
COMMENT This sequence comes from Fig. 1.
FEATURES Region: rfp operon.

	Location/Qualifiers
source	1. .2259
	/organism="Shigella dysenteriae"
	/mol_type="genomic DNA"
	/serotype="1"
	/db_xref="taxon:622"
	/clone="pGom2-1"
	/plasmid="pHW400"
	/note="W30864"
gene	236. .1045
	/gene="rfpA"
CDS	236. .1045
	/gene="rfpA"
	/note="25 kda polypeptide"
	/codon_start=1
	/transl_table=11
	/product="RfpA"
	/protein_id="AAC60479.1"
	/db_xref="GI:688321"
	/translation="MKNHKVSIIIPCFNNGKTIERCVLSALKQTHRNIEIICINDGPS DNSWNILKKLSSLYGNVFAFNNEDNSGPSFSRIKGVSLSTGHFLSFLDADDYWHPKKL ELQLSFINDENLDFLGSTCSIGEKNNQEINQGIKKEHLKLKIISFNMMLFKNYFQTPA VIMKRDIFIPFNENQRFSEDYMSWLVIVYNKKTNVDLIYGRDLVFLDKFNFGVSGLSG NLWLMEKWELKNIFNFLLKGKIMAVPAILFSLIKYERRCALTKKNKGKGNK"
gene	1045. .2178
	/gene="rfpB"
	/note="RfpB"
CDS	1045. .2178
	/gene="rfpB"
	/note="40 kDa membrane-bound galactosyl-transferase; conceptual translation presented here differs from translation in publication"
	/codon_start=1
	/transl_table=11
	/product="RfpB"
	/protein_id="AAC60480.2"
	/db_xref="GI:18391088"
	/translation="MKISIIGNTANAMILFRLDLIKTLTNKGISVYAFATDYNDSSKE IIKKAGAIPVDYNLSRSGINLAGDLWNTYLLSKKLKKIKPDAILSFFSKPSIFGSLAG IFSGVKNNNAMLEGLGFLFTEQPHGTPLKTKLLKNIQVLLYKIIFPHINSLILLNKDD YHDLIDKYKIKLKSCHILGGIGLDMNNYCKSTPPTNEISFIFIARLLAEKGVNEFVAA AKKIKKTHPNVEFIILGAIDKENPGGLSESDVDTLIKSGVISYPGFVSNVADWIEKSS VFVLPSYYREGVPRSTQEAMAMGRPILTTNLPGCKETIIDGVNGYVVKKWSHEDLAEK MLKLINNPEKIISMGEESYKLARERFDANVNNVKLLKILGIPD"
CDS	2241. .>2259
	/note="orf3 3' of rfpB"
	/codon_start=1
	/transl_table=11
	/product="unknown"
	/protein_id="AAC60481.1"
	/db_xref="GI:1683315"
	/translation="MTLPSH"

ORIGIN

Query Match 25.4%; Score 303.4; DB 1; Length 2259;
Best Local Similarity 54.4%; Pred. No. 1.9e-36;
Matches 632; Conservative 0; Mismatches 526; Indels 3; Gaps 1;

```
Qy         16 ATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAAAAAATAGTGATCATA 75
Db       1003 ATGCGCTTTAACAAAGAAAAATAAAGGTAAGGGTAATAAATAATGAAGATCTCAATAATA 1062

Qy         76 GGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATCATGAATTTAGTGAGG 135
Db       1063 GGGAACACAGCAAATGCTATGATTTTGTTTAGATTGGATTTAATAAAAACACTAACCAAC 1122

Qy        136 CAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAAGATCTTAAAGTACTT 195
Db       1123 AAAGGGATTTCAGTCTATGCTTTTGCTACTGACTATAATGATTCATCCAAGGAAATAATA 1182

Qy        196 TCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTT 255
Db       1183 AAAAAAGCAGGCGCCATTCCTGTTGATTATAATTTAAGTCGCAGTGGTATTAACCTTGCT 1242

Qy        256 AAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGATATTTCCCCAGATATT 315
Db       1243 GGTGATTTATGGAATACTTACTTATTAAGTAAAAAACTAAAGAAGATAAAACCAGATGCT 1302

Qy        316 GTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCTTCAAAGTTGTCAAAA 375
Db       1303 ATTTTATCTTTTTTTTCAAAGCCCTCTATCTTTGGATCGTTGGCTGGTATTTTTTCAGGC 1362

Qy        376 GTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTCACTTATTATAAGGGA 435
Db       1363 GTTAAAAATAATAACGCTATGCTTGAGGGGTTAGGTTTTTTATTTACAGAGCAGCCACAT 1422

Qy        436 AAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCA 495
Db       1423 GGAACTCCGTTAAAAACAAAGTTACTTAAAAATATCCAGGTTCTCCTGTATAAAATAATA 1482

Qy        496 TTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGAT 555
Db       1483 TTTCCACATATCAACTCATTAATACTCCTTAACAAGGATGATTATCATGATTTGATAGAT 1542
```

```
Qy        556 CAGTATAATATTAAAGCTAA---GGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAAT 612
Db       1543 AAATACAAAATAAAATTAAAATCTTGCCATATTCTTGGTGGCATTGGTTTAGATATGAAT 1602

Qy        613 GAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTA 672
Db       1603 AATTACTGTAAAAGCACGCCACCAACAAATGAAATATCATTCATTTTTATAGCTCGTTTG 1662

Qy        673 TTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTAT 732
Db       1663 CTAGCAGAAAAAGGAGTCAATGAGTTTGTTGCTGCCGCAAAAAAAAATAAAAAAAACACAT 1722

Qy        733 CCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAA 792
Db       1723 CCCAATGTTGAATTTATTATACTTGGCGCTATAGATAAGGAAAACCCCGGAGGGTTATCT 1782

Qy        793 AAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAA 852
Db       1783 GAATCTGACGTAGATACTTTAATTAAATCAGGAGTTATTTCTTATCCCGGATTTGTTTCT 1842

Qy        853 AATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAA 912
Db       1843 AATGTGGCTGATTGGATTGAAAAATCAAGCGTATTTGTTCTTCCTTCCTATTATCGAGAG 1902

Qy        913 GGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAAT 972
Db       1903 GGAGTTCCTCGTAGTACACAAGAAGCGATGGCTATGGGGAGGCCGATTTTAACTACTAAT 1962

Qy        973 GTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTT 1032
Db       1963 TTACCAGGCTGCAAAGAAACAATTATTGATGGTGTGAATGGATATGTTGTAAAAAAATGG 2022

Qy       1033 GAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAAGTACTC 1092
Db       2023 TCACATGAAGATCTTGCAGAAAAAATGCTGAAGTTAATTAATAATCCTGAAAAAATAATC 2082

Qy       1093 GAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCTTTTGAAAAAAAT 1152
Db       2083 AGTATGGGAGAAGAAAGTTATAAGTTAGCAAGAGAAAGATTCGATGCAAATGTAAATAAT 2142

Qy       1153 AATAGACTAGCATCAATAATA 1173
Db       2143 GTAAAGTTATTAAAAATACTA 2163
```

```
RESULT 7
SHFRFPA
LOCUS       SHFRFPA                 1776 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  Shigella dysenteriae Rfp protein (rfp) gene, complete cds.
ACCESSION   M96064
VERSION     M96064.1  GI:152782
KEYWORDS    Rfp protein; rfp gene.
SOURCE      Shigella dysenteriae
  ORGANISM  Shigella dysenteriae
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 1776)
  AUTHORS   Klena,J.D., Ashford,R.S. II and Schnaitman,C.A.
  TITLE     Role of Escherichia coli K-12 rfa genes and the rfp gene of
            Shigella dysenteriae 1 in generation of lipopolysaccharide core
            heterogeneity and attachment of O antigen
  JOURNAL   J. Bacteriol. 174 (22), 7297-7307 (1992)
   PUBMED   1385388
COMMENT     Original source text: Shigella dysenteriae (strain 1) DNA.
FEATURES             Location/Qualifiers
     source          1. .1776
                     /organism="Shigella dysenteriae"
                     /mol_type="genomic DNA"
                     /strain="1"
                     /db_xref="taxon:622"
     gene            557. .1690
                     /gene="rfp"
     CDS             557. .1690
                     /gene="rfp"
                     /function="Rfa core modification gene"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="Rfp"
                     /protein_id="AAA26537.1"
                     /db_xref="GI:152783"
                     /translation="MKISIIGNTANAMILFRLDLIKTLTKKGISVYAFATDYNDSSKE
                     IIKKAGAIPVDYNLSRSGINLAGDLWNTYLLSKKLKKIKPDAILSFFSKPSIFGSLAG
                     IFSGVKNNTAMLEGLGFLFTEQPHGTPLKTKLLKNIQVLLYKIIFPHINSLILLNKDD
                     YHDLIDKYKIKLKSCHILGGIGLDMNNYCKSTPPTNEISFIFIARLLAEKGVNEFVLA
                     AKKIKKTHPNVEFIILGAIDKENPGGLSESDVDTLIKSGVISYPGFVSNVADWIEKSS
                     VFVLPSYYREGVPRSTQEAMAMGRPILTTNLPGCKETIIDGVNGYVVKKWSHEDLAEK
                     MLKLINNPEKIISMGEESYKLARERFDANVNNVKLLKILGIPD"
ORIGIN

  Query Match             25.3%;  Score 301.8;  DB 1;  Length 1776;
  Best Local Similarity   54.3%;  Pred. No. 3.6e-36;
  Matches  631;  Conservative    0;  Mismatches  527;  Indels    3;  Gaps    1;

Qy         16 ATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAAAAAATAGTGATCATA 75
Db        515 ATGCGCTTTAACAAAGAAAAATAAAGGTAAGGGTAATAAATAATGAAGATCTCAATAATA 574

Qy         76 GGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATCATGAATTTAGTGAGG 135
Db        575 GGGAACACAGCAAATGCTATGATTTTGTTTAGATTGGATTTAATAAAAACACTAACCAAG 634

Qy        136 CAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAAGATCTTAAAGTACTT 195
Db        635 AAAGGGATTTCAGTCTATGCTTTTGCTACTGACTATAATGATTCATCCAAGGAAATAATA 694

Qy        196 TCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTT 255
Db        695 AAAAAAGCAGGCGCCATTCCTGTTGATTATAATTTAAGTCGCAGTGGTATTAACCTTGCT 754

Qy        256 AAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGATATTTCCCCAGATATT 315
Db        755 GGTGATTTATGGAATACTTACTTATTAAGTAAAAAACTAAAGAAGATAAAACCAGATGCT 814

Qy        316 GTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCTTCAAAGTTGTCAAAA 375
Db        815 ATTTTATCTTTTTTTTCAAAGCCCTCTATCTTTGGATCGTTGGCTGGTATTTTTTCAGGC 874

Qy        376 GTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTCACTTATTATAAGGGA 435
Db        875 GTTAAAAATAATACAGCTATGCTTGAGGGGTTAGGTTTTTTATTTACAGAGCAGCCACAT 934

Qy        436 AAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCA 495
Db        935 GGAACTCCGTTAAAAACAAAGTTACTTAAAAATATCCAGGTTCTCCTGTATAAAATAATA 994

Qy        496 TTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGAT 555
Db        995 TTTCCACATATCAACTCATTAATACTCCTTAACAAGGATGATTATCATGATTTGATAGAT 1054

Qy        556 CAGTATAATATTAAAGCTAA---GGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAAT 612
Db       1055 AAATACAAAATAAAATTAAAATCTTGCCATATTCTTGGTGGCATTGGTTTAGATATGAAT 1114

Qy        613 GAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTA 672
Db       1115 AATTACTGTAAAAGCACGCCACCAACAAATGAAATATCATTCATTTTTATAGCTCGTTTG 1174

Qy        673 TTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTAT 732
Db       1175 CTAGCAGAAAAAGGAGTCAATGAGTTTGTGCTTGCCGCAAAAAAAATAAAAAAAACACAT 1234

Qy        733 CCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAA 792
Db       1235 CCCAATGTTGAATTTATTATACTTGGCGCTATAGATAAGGAAAACCCCGGAGGGTTATCT 1294
```

```
Qy        793 AAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAA 852
Db       1295 GAATCTGACGTAGATACTTTAATTAAATCAGGAGTTATTTCTTATCCCGGATTTGTTTCT 1354

Qy        853 AATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAA 912
Db       1355 AATGTGGCTGATTGGATTGAAAAATCAAGCGTATTTGTTCTTCCTTCCTATTATCGAGAG 1414

Qy        913 GGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAAT 972
Db       1415 GGAGTTCCTCGTAGTACACAAGAAGCGATGGCTATGGGGAGGCCGATTTTAACTACTAAT 1474

Qy        973 GTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTT 1032
Db       1475 TTACCAGGCTGCAAAGAAACAATTATTGATGGTGTGAATGGATATGTTGTAAAAAAATGG 1534

Qy       1033 GAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAAGTACTC 1092
Db       1535 TCACATGAAGATCTTGCAGAAAAAATGCTGAAGTTAATTAATAATCCTGAAAAAATAATC 1594

Qy       1093 GAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCTTTTGAAAAAAAT 1152
Db       1595 AGTATGGGAGAAGAAAGTTATAAGTTAGCAAGAGAAAGATTCGATGCAAATGTAAATAAT 1654

Qy       1153 AATAGACTAGCATCAATAATA 1173
Db       1655 GTAAAGTTATTAAAAATACTA 1675
```

RESULT 8
AY532664
LOCUS AY532664 12033 bp DNA linear BCT 30-MAR-2004
DEFINITION Escherichia coli serotype O103 O-antigen gene cluster, complete sequence.
ACCESSION AY532664
VERSION AY532664.1 GI:45644916
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 12033)
AUTHORS Fratamico,P.M., Briggs,C.E., Debroy,C., Strobaugh,T.P. Jr. and Chen,C.-Y.
TITLE DNA Sequence of the Escherichia coli O103 O Antigen Gene Cluster and Detection of Enterohemorrhagic E. coli O103 by PCR Amplification of the wzx and wzy Genes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 12033)
AUTHORS Fratamico,P.M., Briggs,C.E., Debroy,C., Strobaugh,T.P. Jr. and Chen,C.-Y.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-2004) Eastern Regional Research Center, Agriculture Research Service, U. S. Department of Agriculture, 600 E. Mermaid Ln., Wyndmoor, PA 19038, USA

FEATURES Location/Qualifiers

```
     source          1. .12033
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /serotype="O103"
                     /db_xref="taxon:562"
     gene            107. .1183
                     /gene="rmlB"
     CDS             107. .1183
                     /gene="rmlB"
                     /note="RmlB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dTDP-glucose 4,6-dehydratase"
                     /protein_id="AAS73163.1"
                     /db_xref="GI:45644917"
                     /translation="MKILVTGGAGFIGSAVVRHIINNTQDSVVNVDKLTYAGNLESLA
                     DVSDSERYVFEHADICDSVAMARIFAQHQPDAVMHLAAESHVDRSITGPAAFIETNIV
                     GTYVLLEAARNYWSALDGDKKNSFRFHHISTDEVYGDLPHPDEVNNTEELPLFTETTA
                     YAPSSPYSASKASSDHLVRAWKRTYGLPTIVTNCSNNYGPYHFPEKLIPLVILNALEG
                     KGLPIYGKGDQIRDWLYVEDHARALYTVVTEGKAGETYNIGGHNEKKNIDVVLTICDL
                     LDEIVPKEKSYREQITYVADRPGHDRRYAIDAEKIGRELGWKPQETFESGIRKTINWY
                     LNNNDWCRRVQDGSYQRERLGVIR"
     gene            1180. .2052
     CDS             1180. .2052
                     /gene="rmlA"
                     /note="RmlA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative D-glucose-1-phosphate thymidylyltransferase"
                     /protein_id="AAS73164.1"
                     /db_xref="GI:45644918"
                     /translation="MKGIILAGGSGTRLYPITMGVSKQLLPIYDKPMIYYPLSVLMLA
                     GIQDILIITTPEDQQGFMRLLGDGSQFGINLTYAIQPSPDGLAQAFIIGENFIGNDTV
                     CLMLGDNIFFGQGFTPKLRDAAERTSGATVFGYQVMDPERFGVVEFDRNFKALSIEEK
                     PANPKSNWAVTGLYFYDNNVIDIAKCIKPSKRGELEITAVNEVYLRNNNLTVELLGRG
                     FAWLDTGTHDSLIEAGSFVETVQKRQGMMVACPEEIAWRNGWLNHDELQAKGEFLKKN
                     HYGQYLLRLIKSHE"
     gene            2045. .2452
     CDS             2045. .2452
                     /gene="wbtA"
                     /note="WbtA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dTDP-6-deoxy-3,4-keto-hexulose isomerase"
                     /protein_id="AAS73165.1"
                     /db_xref="GI:45644919"
                     /translation="MNNNVKLIHFDKKGDNRGFLVVVEQNKDIPIEIKRVYYMFDTGS
                     DVRRGYHAHKKLTQIAIPVKGSCKFLLDDGKSVEHVTLDSPSVGLIIEPMIWHEMYDY
                     SSDCVLMVLADDFYDESDYIRNYADFIREMTNA"
     gene            2445. .2978
     CDS             2445. .2978
                     /gene="wbtB"
                     /note="WbtB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative butyryltransferase"
                     /protein_id="AAS73166.1"
                     /db_xref="GI:45644920"
                     /translation="MRNLSLFTKNTKLIPVEINDAEFIYSLRVDSRLNKHISTINGTV
                     IEQAEWIEDYKKREENNEEFYFIIKRTDNDEKIGTIRLYNITKDDGFCWGSWILNDDK
                     MKSSAIESAYLIYKFGFLHKNYKYSYFQVDKKNKEVIAFHKKTGATIIKEDDVNFHPE
                     FYVENFIRLREKYSKYF"
     gene            2975. .4096
     CDS             2975. .4096
                     /gene="wbtC"
                     /note="WbtC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative aminotransferase"
                     /protein_id="AAS73167.1"
                     /db_xref="GI:45644921"
                     /translation="MRGKKMIDFLNLKRINAQYEMELKDACSRVIDSGWYIMGTELKT
                     FEEKFANWCGVKHVIGVANGLDALILVLRAWKEMGKLNDGDEIIVPANTYIASVLAIT
                     ENNLVPVLTEPDPLTFNLSAGGIRKVITGKTKAILPVHLYGQISPMAEIMEIAAEFGL
                     LVLEDCAQAHGSSIDGRKAGSWGHAAGFSFYPGKNLGALGDAGAITTSDDQLADILTA
                     IRNYGSHEKYKNKYQGLNSRLDEMQAAFLNVKLKYLDRENKLRRAIAKKYIEEIKNPL
                     ISLPVIAGNEMSHVWHLFVIRTSERDSLQKHLKTEGITTLVHYPIAPHNQEAYKEYSF
                     LSLPLTERIHREVLSLPMDPTMSIDDINMVIEAINRWKS"
     gene            4084. .5349
     CDS             4084. .5349
                     /gene="wzx"
                     /note="Wzx"
                     /codon_start=1
```

```
                /transl_table=11
                /product="putative O-antigen flippase"
                /protein_id="AAS73168.1"
                /db_xref="GI:45644922"
                /translation="MEKLNFAKVTILSGVVTCFRLICGLLISKVVAIYTGPIGIANLG
                QLQNFVAFINGFISSQVSQGVNRYSAENQGDYENACFYWRASIKLSFIACCFIISTGV
                LFSSKISEYLFYQSDLYWLIIIALCALPLNIVNNIFLGVLNGLNEYNRFFLANVLAIS
                SSLLSMVGLVYFFGLKGALVSASLNNAVAGVWLITIIIKRPWFKFKYWVGHTPRHNIT
                QMKNYFYMGVIGALTGPISMIVVRTILTNNFSLEDAGYWQAVNRISEAYLAVLTTALT
                VYYFPKTAAARRYSEYITLLKTGACIVVPLALSMALTIYGLKDFIISILFTADFIRAR
                ELFLFQNIGDFLRICSWLFATILLAKGYFKINAILEVMFSIIFPVLVKFLIINFGLNA
                VSIAYSITYALYFIVVCIIYCWHLTKKAL"
gene            5353. .6339
                /gene="wbtD"
CDS             5353. .6339
                /gene="wbtD"
                /note="WbtD"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyltransferase"
                /protein_id="AAS73169.1"
                /db_xref="GI:45644923"
                /translation="MIISSDNLSVIIPVYNEEKNVINILQSLECQSLTGFNVIVIDDG
                SKDMTASLVKEYKPSSYSLSLIQQSNMGAARARENAINFTHSEYIAFIDSDDSLSSDA
                LEKALAPMLDNKDIDISLFELVHIKNLNHDTNNIFVPYSTTKLICGEEAFANCISYWG
                LHGFGIYKRKLIQKSYGIYYKYNKNKENYINNDEVISRISFGLSKNIYLSSGRYFFVQ
                NMDSTTRRINESYYKVINNAVYLREYIDEEIKYNDFDCLGEANKLLVSTIWGVFVRYQ
                KWKKKFSDETNGKWREAIKKGMQYIKKINNNKNINLHIKSKMQLYIISKLVS"
gene            6342. .7232
                /gene="wbtE"
CDS             6342. .7232
                /gene="wbtE"
                /note="WbtE"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyltransferase"
                /protein_id="AAS73170.1"
                /db_xref="GI:45644924"
                /translation="MKNLLITIIVASYNSKDYILETLESCINQQYQNIEIIIVDDCSS
                DNSVELIKQWCKEKKSTSPKIRCILVESERNKGIPANLNNALPFINGDWIKCIGSDDI
                LLPEAISEFVNRLEKCSNLDNIGAVFTYFQTFGYSVNVSTRYPSSWTRTICQMPPSWL
                KKAAAMIHFNNVAPCAFINKNHFSGFDTSYKLLEDLPLWLKMIKNNKNTLFFDYTTVL
                YRLHDSQITSLNNIKVNDVLINDLCRLNYYRKNNGYYLAYLHHKFNLYCAVKRMRFYR
                YLKIINPLNLLIVLYERVKK"
gene            7232. .8380
                /gene="wzy"
CDS             7232. .8380
                /gene="wzy"
                /note="Wzy"
                /codon_start=1
                /transl_table=11
                /product="putative O-antigen polymerase"
                /protein_id="AAS73171.1"
                /db_xref="GI:45644925"
                /translation="MYILSMAIPLFFISLCWHRFSHQTQVAFRLLFIVFFSLSFFSYY
                TNGVDWSVYYLRFIEDEGLFTSFEFGFVIILKLLLIVSDDNFGLAILLYYFLCFVLLS
                LILKKYKVNEPLFLGCLLLLFGYNLFLEQLRQLLACIIVFYAMLLYNQNKNLIELCIL
                VITASTFHVSAVIILPTIFLISFRNVTTFIIITVSSITSIVVFMIALYSAISSLAAMS
                FVFAKIDYYLHQYPVVLNFGWLNILDALFILFYMRYRNAIDRQININILTRIIFVGSV
                IHLFSGSITFLARVCFYFYFVGVYIYCYSLNDKTPRVFAIKNFNTLVLAAFVNVVLLL
                NFASYFRNEQAPVGFYNMNFQFFNTFNDNYLRTNAYDKYFKGMQGSVD"
gene            8383. .9474
                /gene="wbtF"
CDS             8383. .9474
                /gene="wbtF"
                /note="WbtF"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyltransferase"
                /protein_id="AAS73172.1"
                /db_xref="GI:45644926"
                /translation="MSKRIIFVITGLGLGGAEKQICLLADKLYLHNNEVNIVVLNGNV
                NVTPLNENIRIHNLHMKKNPLGLLNAVLRLVKIISYFKPDIVHGHMFHANIVARLAKV
                FTYNKYKLICTAHSKNEGGHIRMLVYRLTDFLCDATTNVSNEALDVFIKKKAFSKSKS
                IPVYNGIDTDLFKYDPLQRDKIRHSLNIINDDSLILSVGRLTDAKDYPNLLKALLELP
                ERYKLVIIGEGEARNEVERTILNYGLEKRVKLLGSIANVATYYSACDIYVSSSKWEGF
                GLVVAEAMSCQRLVVATNAGGVAEVVGDMNYIVPVSDPQLLANKINEVMEFDHETKME
                VMYRNREFVLKMFSIDMIVNQWMKLYSSI"
gene            9660. .10781
                /gene="wbtG"
CDS             9660. .10781
                /gene="wbtG"
                /note="WbtG"
                /codon_start=1
                /transl_table=11

Query Match          21.2%;  Score 252.6;  DB 1;  Length 12033;
Best Local Similarity  53.0%;  Pred. No. 4.4e-29;
Matches  612;  Conservative    0;  Mismatches  534;  Indels    9;  Gaps    3;

Qy        30 ATATTATTTACTTTTAGAATGTGATATGAAAAAAATAGTGATCATAGGCAATGTAGCGTC 89
Db      9632 ATACTGTTATATTGTAAGGTTTTTAAATATGAAAGTGCTTTTAATAGGTAACCAATCTAG 9691
Qy        90 AATGATGTTAAGGTTCAGGAAAGAATTAATCATGAATTTAGTGAGGCAAGGTGATAATGT 149
Db      9692 TACTATTATTTTGTTCAGAAAGGCTGTTATTGAAAATTTGATATCAAAAAATATTAAAGT 9751
Qy       150 ATATTGTCTAGCAAATGATTTTTCCACTGAAGATCTTAAAGTACTTTCGTCATGGGGCGT 209
Db      9752 ATATACACTTACGATGGATAATAATGAGGGTAATTTTAATAAAATTAAAAAAATAGGTGC 9811
Qy       210 TAAGGGGGTTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTTAAGGATATAATTGC 269
Db      9812 AATCCCTTTATCCTATTCTTTTTCACGATCTGGTATTAATCCTATATCTGATTTTTTTAA 9871
Qy       270 TGTTTATGAACTAAAAAAAATTCTTAAGGATATTTCCCCAGATATTGTATTTTCATATTT 329
Db      9872 TACTTTTTTTTTGGCAAGGAAAATAAAAGAGATTAATCCTGATTCTGTTCTTTGCTTTTT 9931
Qy       330 TGTAAAGCCAGTAATATTTGGAACTATTGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGT 389
Db      9932 CCCAAAACCAGTCATATATGGTACTCTTGCCGCTATGATTGCTGGCGTAAAAAAACGTTA 9991
Qy       390 TGGAATGATTGAAGGTCTAGGTAATGCCTTCACTTATTATAAGGGAAAGCAGACCACAAA 449
Db      9992 CGCTTTATTGGAAGGGTTGGGTTTTTGTTTCACACAAAATAAAGATAAAGACTCGTTTAA 10051
Qy       450 AACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGCTTGA 509
Db     10052 AAAAATAGTCTTAAGAAAGATTCAAGGTTTTTTATATAAGATGACATTACCTTTTGCAAC 10111
Qy       510 TGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAA 569
Db     10112 CAAAGTTCTTTTTCTCAATAATGATGACTATCATGACTTGGTTGTTAAAAACAATATTAA 10171
Qy       570 AGCTAAGG---TAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAA 626
Db     10172 AATCAAAGATTATAAAGTTATTGGTGGCATAGGTGTAAATCTCAATGAATTTAGTTATCA 10231
Qy       627 AGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGG 686
Db     10232 TCCACCAATTTGTGGGGAGATCCATTTTGGTATGGTGGCAAGATTATTGAGAGAGAAGGG 10291
Qy       687 GATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATT 746
Db     10292 CGTATGTGAATTTGTTGAAGCGGCTAAAATAATAAAGAAAAAATATCCAATGGTCCGATT 10351
Qy       747 TGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGA 806
Db     10352 TAGTATAGCTGGTGATTTTGA---TGATAATCCGGGTGGAATTGATGCTGCACAAATTGT 10408
Qy       807 ATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTG 866
Db     10409 AAATTGGAAAGCGGAAGGTTATGTTGAATTTTTAGGACCGTGCTCTTATATAAAAGAATA 10468
```

```
Qy        867 GTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGT 926
              ||      | ||      || ||||| ||||||   || ||||||||||| || || |
Db      10469 TTTGAGTGATATCAGCGTCTTTGTATTACCT---TCTTATCGAGAAGGTGTCCCCAAAAG 10525

Qy        927 GATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAG 986
              |  || ||||| ||||| || ||| |   |  | || || ||  |||| || || ||| |
Db      10526 TACTCAGGAAGCAATGGCAATAGGTCGTGCGATTATTACGACCGATGTGCCAGGTTGTCG 10585

Qy        987 GGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACT 1046
              || | | ||| ||||||     ||||| | | |  ||||||||  || |    |  |||
Db      10586 AGAAACAGTAATTGATGGATATAATGGATATCTTGTACCTCCATGGGATACACAAAAACT 10645

Qy       1047 GGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGC 1106
              ||    |   ||| ||  |||||| |  |    ||      |  ||   ||||| |
Db      10646 TGCTTTAGTTATGGAAACATTTATTCATGAACGTGAGCTTATTGTCACTATGGGATTAAA 10705

Qy       1107 TGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCTTTTGAAAAAAATAATAGACTAGCATC 1166
              | |  | | ||| |||     ||| ||| |||    ||||       |||   |||
Db      10706 TAGTCGAAGGTTCGCATTGTCAAAATTTAATGAAAGTGAATCTGCAAATTCTTTAGTTAA 10765

Qy       1167 AATAATAAAATCAAA 1181
              || ||||   || |
Db      10766 TATTATAACTGCATA 10780
```

```
RESULT 9
NMA1Z2491
LOCUS       NMA1Z2491              340806 bp    DNA     linear   BCT 16-APR-2005
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 1/7.
ACCESSION   AL162752 AL157959
VERSION     AL162752.2  GI:7378778
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Neisseria.
REFERENCE   1  (bases 1 to 340806)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  PUBMED    10761919
REFERENCE   2  (bases 1 to 340806)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .340806
                     /organism="Neisseria meningitidis Z2491"
                     /mol_type="genomic DNA"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     gene            complement(38. .358)
                     /gene="NMA0001"
     CDS             complement(38. .358)
                     /gene="NMA0001"
                     /note="NMA0001, unknown, len: 106 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0001"
                     /protein_id="CAB83321.1"
                     /db_xref="GI:7378779"
                     /db_xref="InterPro:IPR007113"
                     /db_xref="UniProt/TrEMBL:Q9JX93"
                     /translation="MSNTQTIRSADFTTSRAWGALDIANMNGTTVRLHWTDQPYKWHV
                     NDGEEVFAVMDGEVDMHYRENGEEHIVRLKSGDIFYAGIGTEHVAHPRGETRILVIEK
                     EGSV"
     gene            complement(414. .2438)
                     /gene="nuoL"
     CDS             complement(414. .2438)
                     /gene="nuoL"
                     /EC_number="1.6.5.3"
                     /note="NMA0002, nuoL, NADH dehydrogenase I chain L, len:
                     674 aa; similar to many e.g. NUOL_RHOCA NADH dehydrogenase
                     I chain L (EC 1.6.5.3) (700 aa). Contains Pfam match to
                     entry PF00361 oxidored_q1, NADH-Ubiquinone/plastoquinone
                     (complex I), various chains, and to entry PF00662
                     oxidored_q1_N, NADH-Ubiquinone oxidoreductase (complex I),
                     chain 5 N-terminus"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase I chain L"
                     /protein_id="CAB83322.1"
                     /db_xref="GI:7378780"
                     /db_xref="GOA:Q9JX92"
                     /db_xref="InterPro:IPR001516"
                     /db_xref="InterPro:IPR001750"
                     /db_xref="InterPro:IPR003916"
                     /db_xref="InterPro:IPR003945"
                     /db_xref="UniProt/Swiss-Prot:Q9JX92"
                     /translation="MNDMTLYLIIALVPLAGSLIAGLFGNKIGRAGAHTVTILGVAVS
                     AVLSAYVLWGFLNGSRAKFDENVYTWLTMGGLDFSVGFLVDTMTAMMMVVVTGVSLMV
                     HIYTIGYMHDEKVGYQRFFSYISLFTFSMLMLIMSNNFIQLFFGWEAVGLVSYLLIGF
                     YFKRPSATFANLKAFLINRVGDFGFLLGIGLVLAYFGGSLRYQDVFAYLPNVQNATIQ
                     LFPGVEWSLITVTCLLLFVGAMGKSAQFPLHVWLPDSMEGPTPISALIHAATMVTAGL
                     FMVSRMSPIYEMSSTALSVIMVIGAITALFMGFLGVIQNDIKRVVAYSTLSQLGYMTV
                     ALGASAYSVAMFHVMTHAFFKALLFLAAGSAIIGMHHDQDMRHMGNLKKYMPITWLTM
                     LIGNLSLIGTPFFSGFYSKDSIIEAAKYSTLPGSGFAYFAVLASVFVTAFYAFRQYFM
                     VFHGEEKWRSLPEHHSDGHGEEHHGLGKNDNPHESPLVVTLPLILLAVPSVIIGYIAI
                     EPMLYGDFFKDVIFVNADAHPTMHIMKEEFHGALAMVSHSLHSPVLYLAIAGVLSAWL
                     LYVKLPHLPAKIAQAFRPVYVLFENKYYLDALYFNVFAKGTRALGTFFWKVGDTAIID
                     NGIVNGSARLVGAVAAQVRKVQTGFIYTYAAAMVFGVLVLLGMTFWGLFR"
     misc_feature    1035. .1044
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(1116. .2030)
                     /gene="nuoL"
                     /note="Pfam match to entry PF00361 oxidored_q1,
                     NADH-Ubiquinone/plastoquinone (complex I), various chains,
                     score 351.80, E-value 7.3e-102"
     misc_feature    complement(2061. .2243)
                     /gene="nuoL"
                     /note="Pfam match to entry PF00662 oxidored_q1_N,
                     NADH-Ubiquinone oxidoreductase (complex I), chain 5
                     N-terminus, score 77.20, E-value 3.5e-19"
     gene            complement(2487. .2828)
                     /gene="NMA0003"
     CDS             complement(2487. .2828)
                     /gene="NMA0003"
                     /note="NMA0003, unknown, len: 113 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0003"
                     /protein_id="CAB83323.1"
                     /db_xref="GI:7378781"
                     /db_xref="UniProt/TrEMBL:Q9JX91"
                     /translation="MMDKNQLEQEFHKAMLNIYQEALNLPQPYKATRFLQIVNEFGGK
                     EAADKLLSTGEKKTQTGFTELILSGGGVHALKYSMEYLVLQKPWCDLFTEEQLAVARK
                     RLERVGFVSPK"
     RBS             complement(2833. .2837)
     gene            complement(2952. .3527)
                     /gene="NMA0004"
     CDS             complement(2952. .3527)
```

```
                     /gene="NMA0004"
                     /note="NMA0004, unknown, len: 191 aa; similar to
                     hypothetical proteins e.g. Y977_HAEIN HI0977 (191 aa),
                     fasta scores; E(): 0; 80.5% identity in 190 aa overlap,
                     and to proteins involved in cell cycle e.g. TR:O54679
                     (EMBL:AF036487) Lactococcus lactis Plasmid pNZ4000
                     putative mobilization protein (200 aa), fasta scores; E():
                     0, 55.1% identity in 167 aa overlap, and FIC_ECOLI cell
                     filamentation protein FIC (200 aa), fasta scores; E():
                     0.093, 29.4% identity in 126 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein NMA0004"
                     /protein_id="CAB83324.1"
                     /db_xref="GI:7378782"
                     /db_xref="GOA:Q9JQR9"
                     /db_xref="InterPro:IPR000102"
                     /db_xref="InterPro:IPR003812"
                     /db_xref="UniProt/TrEMBL:Q9JQR9"
                     /translation="MPSENPIGKTMKSIDEQSLHNARRLFESGDIDRIEVGTTAGLQQ
                     IHRYLFGGLYDFAGQIREDNISKGGFRFANAMYLKEALVKIEQMPERTFEEIIAKYVE
                     MNIAHPFLEGNGRSTRIWLDLVLKKNLKKVVNWQNVSKTLYLQAMERSPVNDLELRFL
                     LKDNLTDDVDNREIIFKGIEQSYYYEGYEKG"
misc_feature         complement(3516. .3525)
                     /gene="NMA0004"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
gene                 complement(3561. .3866)
                     /gene="nuoK"
CDS                  complement(3561. .3866)
                     /gene="nuoK"
                     /EC_number="1.6.5.3"
                     /note="NMA0005, nuoK, NADH dehydrogenase I chain K, len:
                     101 aa; simlar to many e.g. NUOK_RHOCA NADH dehydrogenase
                     I chain K (EC 1.6.5.3) (102 aa), fasta scores; E():
                     1.1e-24; 65.0% identity in 100 aa overlap. Contains Pfam
                     match to entry PF00420 oxidored_q2,
                     NADH-ubiquinone/plastoquinone oxidoreductase chain 4L"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase I chain K"
                     /protein_id="CAB83325.1"
                     /db_xref="GI:7378783"
                     /db_xref="GOA:Q9JQU9"
                     /db_xref="InterPro:IPR001133"
                     /db_xref="InterPro:IPR003214"
                     /db_xref="InterPro:IPR003215"
                     /db_xref="UniProt/TrEMBL:Q9JQU9"
                     /translation="MITLTHYLVLGALLFGISAMGIFMNRKNVLVLLMSIELMLLAVN
                     FNFIAFSQHLGDTAGQIFVFFVLTVAAAESAIGLAIMVLVYRNRQTINVADLDELKG"
misc_feature         complement(3564. .3863)
                     /gene="nuoK"
                     /note="Pfam match to entry PF00420 oxidored_q2,
                     NADH-ubiquinone/plastoquinone oxidoreductase chain 4L,
                     score 141.50, E-value 1.5e-38"
gene                 complement(3863. .4534)
                     /gene="nuoJ"
CDS                  complement(3863. .4534)
                     /gene="nuoJ"
                     /EC_number="1.6.5.3"
                     /note="NMA0006, nuoJ, NADH dehydrogenase I chain J, len:
                     223 aa; similar to many e.g. NUOJ_RHOCA NADH dehydrogenase
                     I chain J (EC 1.6.5.3) (202 aa), fatsa scores; E():
                     9.7e-21; 41.3% identity in 201 aa overlap. Contains Pfam
                     match to entry PF00499 oxidored_q3,
                     NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase I chain J"
                     /protein_id="CAB83326.1"
                     /db_xref="GI:7378784"
                     /db_xref="GOA:Q9JX90"
                     /db_xref="InterPro:IPR001457"
                     /db_xref="UniProt/TrEMBL:Q9JX90"
                     /translation="MTFQLILFYIFAVIILYGAIKTVTAKNPVHAALHLVLTFCVSAM
                     LWMLMQAEFLGVTLVVVYVGAVMVLFLFVVMMLNIDIEEMRAGFWRHAPVAGVVGTLL
                     AVALILILVNPKTDLAAFGLMKDIPADYNNIRDLGSRIYTDYLLPFELAAVLLLLGMV
                     AAIALVHRKTVNPKRMDPADQVKVRADQGRMRLVKMEAVKPQTESAEESEVSDDLKPK
                     EEGKA"
RBS                  complement(3874. .3879)
                     /gene="nuoJ"
```

Query Match 18.7%; Score 223.8; DB 1; Length 340806;
Best Local Similarity 52.1%; Pred. No. 2.7e-25;
Matches 585; Conservative 0; Mismatches 517; Indels 21; Gaps 3;

```
Qy      32 ATTATTTACTTTTAGAATGTGATATGAAAAAAATAGTGATCATAGGCAATGTAGCGTCAA 91
Db   39691 ATCCTATAATTCAAGGTGTCCGCCTTATGAAAATCGTTTTTATCACAACAGTCGCATCCA 39750

Qy      92 TGATGTTAAGGTTCAGGAAAGAATTAATCATGAATTTAGTGAGGCAAGGTGATAATGTAT 151
Db   39751 GCATTTACGGTTTCCGCGCCCCCGTCATTAAAAAATTAATCGGCAAAAACCATCAGGTGT 39810

Qy     152 ATTGTCTAGCAAATGATTTTTCCACTGAAGATCTTAAAGTACTTTCGTCATGGGGCGTTA 211
Db   39811 ATGCCTTTGTATCGGAGTTTTCCGACAACGAATTGGACATTATCAGGGAAATGGGGGTAA 39870

Qy     212 AGGGGGTTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTTAAGGATATAATTGCTG 271
Db   39871 CACCCGTTACCTACCGTTCAAACCGCAGCGGGCTGAACCCGTTTTCCGATATAAAATCCA 39930

Qy     272 TTTATGAACTAAAAAAAATTCTTAAGGATATTTCCCCAGATATTGTATTTTCATATTTTG 331
Db   39931 CCTTCCTCATCTTTAAAGAACTCAAAAAAATATCGCCGGATTTGGTTTTCCCTTATTTCG 39990

Qy     332 TAAAGCCAGTAATATTTGGAACTATTGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGTTG 391
Db   39991 CAAAACCCGTGATTTTCGGCACTTTTGCTGCAAAATTGGCAGGCGTGCCCAGAATCGTCG 40050

Qy     392 GAATGATTGAAGGTCTAGGTAATGCCTTCACTTATTATAAGGGAAAGCAGACCACAAAAA 451
Db   40051 GGATGCTGGAAGGTTTGGGATTCGCATTTACCCCGCAGCCGGAAGGCATACCGTTAAAAA 40110

Qy     452 CTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGCTTGATG 511
Db   40111 CAAAAATCATAAAGGGGATTTTAATCACCTTGTACCGCATTGCCCTGCCGATGTTGGAAA 40170

Qy     512 ATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAA-- 569
Db   40171 GCCTGATTGTATTAAACCCCGACGACAAAGACGAACTGACGGACAAATACGGCATCAAAA 40230

Qy     570 -AGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAAG 628
Db   40231 TAAAAAACATCCATATTTTGGGCGGAATCGGTCTGGATTTGCGGCAATATCCTTATTCCG 40290

Qy     629 AGCC------------ACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAA 676
Db   40291 AGGCGGATATTCCCGATGAAAAAGAACCCGTAAAATTCCTCTTTATCGGCAGATTTCTGA 40350

Qy     677 GAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAA 736
Db   40351 AAGAAAAGGGGATTGATGATTTTATTCGGGCGGCGGAACAGGTTAAGGACAAATACCCCG 40410

Qy     737 GTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAA 796
Db   40411 ATACGGTTTTTACCGCTTTGGGCGCAATCG------ACAAATCACGCGGGGGGGGGGGG 40464

Qy     797 ATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAAAATG 856
Db   40465 GCGATTTGGAACGGCTTGCCGCCCGCGATATTATCCGTTTCCCCGGTTTTGTGAACAATG 40524

Qy     857 TTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAAGGCG 916
Db   40525 TTTCCGAAGTGATAAAAGAACATCATATATTCGTATTGCCGTCTTATTATAGGGAAGGCG 40584
```

```
Qy        917 TACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAATGTAC 976
Db      40585 TTCCCCGAAGCACTCAGGAGGCAATGGCCGTCGGCAGGGCAGTGATTACGACGGATGTCC 40644

Qy        977 CTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAA 1036
Db      40645 CCGGATGCAGGGAAACGGTCGCCGACAAGGTCAACGGCTTCCTGATCGAACCTTGGAATC 40704

Qy       1037 TTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAAGTACTCGAAA 1096
Db      40705 CCCGCATCTTGGCCGAAAAAATGATTTATTTTATCGAAAACAGGGAAGCCATCCGCCTGA 40764

Qy       1097 TGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGC 1139
Db      40765 TGGGGAATGCAAGTTATGCGATTGCCAAAGATAAATTCGATGC 40807
```

RESULT 10
AF485419
LOCUS AF485419 1582 bp DNA linear BCT 19-JUL-2002
DEFINITION Neisseria gonorrhoeae strain MS11A pilin glycosyl transferase A (pgtA) gene, pgtA-1 allele, complete cds.
ACCESSION AF485419
VERSION AF485419.1 GI:20219023
KEYWORDS .
SOURCE Neisseria gonorrhoeae
ORGANISM Neisseria gonorrhoeae
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 1582)
AUTHORS Banerjee,A., Wang,R., Supernavage,S.L., Ghosh,S.K., Parker,J., Ganesh,N.F., Wang,P.G., Gulati,S. and Rice,P.A.
TITLE Implications of Phase Variation of a Gene (pgtA) Encoding a Pilin Galactosyl Transferase in Gonococcal Pathogenesis
JOURNAL J. Exp. Med. 196 (2), 147-162 (2002)
PUBMED 12119340
REFERENCE 2 (bases 1 to 1582)
AUTHORS Banerjee,A.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2002) Biology, Catholic Univ. of America, 620 Michigan Avenue, NE, Washington, DC 20064, USA
FEATURES Location/Qualifiers
source 1. .1582
/organism="Neisseria gonorrhoeae"
/mol_type="genomic DNA"
/strain="MS11A"
/db_xref="taxon:485"
gene 455. .1582
/gene="pgtA"
/allele="1"
CDS 455. .1582
/gene="pgtA"
/allele="1"
/note="galactosyl transferase"
/codon_start=1
/transl_table=11
/product="pilin glycosyl transferase A"
/protein_id="AAM15779.1"
/db_xref="GI:20219024"
/translation="MKIVFITTVASSIYGFRAPVIKKLIGKNHQVYAFVSEFSDNELD
IIREMGVTPVTYRSNRSGVNPFSDIKSTFLIFKALKKISPDLVFPYFAKPVIFGTFAA
KLAGVPRIVGMLEGLGFAFTPQPEGIPLKTKIIKGILIALYRITLPMLESLIVLNPDD
KDELLHQYGIKIKNIHILGGIGLDLRQYPYSEADIPDEKEPVKFLFIGRFLKEKGIDD
FIRAAEQVKGKYPDTVFTALGAIDKPRGSGDLEQPAARDIIRFPGFVNNVSEVIKEHH
IFVLPSYYREGVPRSTQEAMAVGRAVITTDVPGCRETVADKVNGFLIEPWNPRILAEK
MIYFIENRAAVRLMANASYAIAKDKFDAEKVDLKFLDILKA"
ORIGIN

Query Match 18.7%; Score 223; DB 1; Length 1582;
Best Local Similarity 52.5%; Pred. No. 2.7e-24;
Matches 577; Conservative 0; Mismatches 495; Indels 27; Gaps 3;

```
Qy         56 TGAAAAAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAAT 115
Db        453 TTATGAAAAATCGTTTTTATCACAACAGTCGCATCCAGCATTTACGGTTTCCGCGCCCCCG 512

Qy        116 TAATCATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCA 175
Db        513 TCATTAAAAAATTAATCGGCAAAAACCATCAGGTGTATGCCTTTGTATCGGAGTTTTCCG 572

Qy        176 CTGAAGATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACT 235
Db        573 ATAATGAGTTGGACATTATCAGGGAAATGGGGGTTACACCCGTTACCTACCGGTCAAACC 632

Qy        236 CAAAGGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTA 295
Db        633 GCAGCGGGGTAAACCCGTTTTCCGATATAAAATCCACCTTCCTCATATTTAAAGCACTCA 692

Qy        296 AGGATATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTA 355
Db        693 AAAAAATATCGCCGGATTTGGTTTTCCCTTATTTCGCAAAACCCGTGATTTTCGGCACTT 752

Qy        356 TTGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATG 415
Db        753 TTGCCGCAAAATTGGCAGGCGTGCCCAGAATCGTCGGGATGCTGGAAGGTTTGGGATTCG 812

Qy        416 CCTTCACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAA 475
Db        813 CATTTACCCCGCAGCCGGAAGGCATACCGTTAAAAACAAAAATCATAAAGGGGATTTTAA 872

Qy        476 TTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATG 535
Db        873 TCGCCTTGTACCGCATTACCCTGCCGATGTTGGAAAGCCTGATTGTATTAAACCCCGACG 932

Qy        536 ATAAAAAAGATTTAATCGATCAGTATAATATTAA---AGCTAAGGTAACAGTGTTAGGTG 592
Db        933 ACAAAGACGAGTTGCTGCATCAATACGGCATCAAAATAAAAAACATCCATATTTTGGGCG 992

Qy        593 GGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCG----------AAAG 640
Db        993 GAATCGGTCTGGATTTGCGGCAATATCCTTATTCCGAGGCGGATATTCCCGATGAAAAAG 1052

Qy        641 AGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTA 700
Db       1053 AACCCGTAAAATTCCTCTTTATCGGCAGATTTCTGAAAGAAAAGGGGATTGATGATTTTA 1112

Qy        701 TTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAG 760
Db       1113 TTCGGGCGGCGGAACAGGTTAAGGGCAAATACCCCGATACGGTTTTTACCGCTTTGGGCG 1172

Qy        761 GTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAG 820
Db       1173 CAATCGACA----------AACCACGCGGGAGCGGGGATTTGGAACAGCCTGCCGCCC 1220

Qy        821 AACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTT 880
Db       1221 GCGATATTATCCGTTTCCCCGGTTTTGTAAACAATGTTTCCGAAGTGATAAAAGAACATC 1280

Qy        881 CTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTA 940
Db       1281 ATATATTCGTATTGCCGTCTTATTATAGGGAAGGCGTTCCCCGAAGCACCCAGGAGGCAA 1340

Qy        941 TGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATG 1000
Db       1341 TGGCCGTCGGCAGGGCGGTGATTACGACGGATGTCCCCGGATGCAGGGAAACGGTTGCCG 1400

Qy       1001 ATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGA 1060
Db       1401 ACAAGGTCAACGGCTTCCTGATCGAACCTTGGAATCCCCGCATCTTGGCCGAAAAAATGA 1460

Qy       1061 AATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTG 1120
Db       1461 TTTATTTTATCGAAAACAGGGCTGCCGTCCGCCTGATGGCGAATGCAAGTTATGCGATTG 1520

Qy       1121 CAGAAAAAAACTTTGATGC 1139
```

```
Db       1521 CCAAAGATAAATTCGATGC 1539


RESULT 11
AE002098_02/c
WPCOMMENT
  Sequence split into 23 fragments  LOCUS AE002098 Accession AE002098
     Fragment Name      Begin     End
     AE002098_00            1    110000
     AE002098_01       100001    210000
     AE002098_02       200001    310000
     AE002098_03       300001    410000
     AE002098_04       400001    510000
     AE002098_05       500001    610000
     AE002098_06       600001    710000
     AE002098_07       700001    810000
     AE002098_08       800001    910000
     AE002098_09       900001   1010000
     AE002098_10      1000001   1110000
     AE002098_11      1100001   1210000
     AE002098_12      1200001   1310000
     AE002098_13      1300001   1410000
     AE002098_14      1400001   1510000
     AE002098_15      1500001   1610000
     AE002098_16      1600001   1710000
     AE002098_17      1700001   1810000
     AE002098_18      1800001   1910000
     AE002098_19      1900001   2010000
     AE002098_20      2000001   2110000
     AE002098_21      2100001   2210000
     AE002098_22      2200001   2272360
  Continuation (3 of 23) of AE002098 from base 200001 (AE002098 Neisseria meningitidis MC5

  Query Match             18.6%;  Score 222.6;  DB 1;  Length 110000;
  Best Local Similarity   52.4%;  Pred. No. 6.2e-25;
  Matches  576;  Conservative    0;  Mismatches  499;  Indels   24;  Gaps    3;

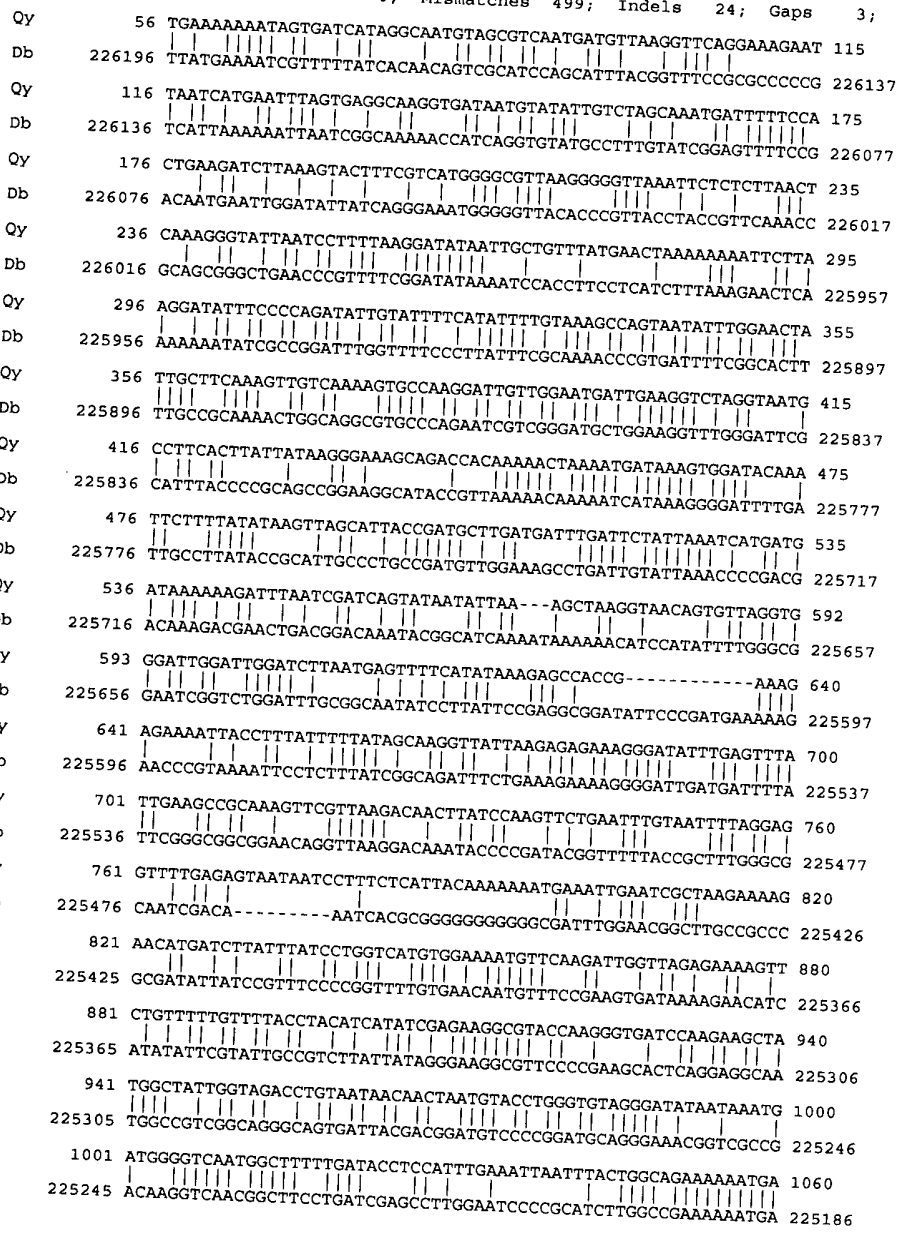
Qy         56 TGAAAAAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAAT 115
Db      26203 TTATGAAAATCGTTTTTATCACAACAGTCGCATCCAGCATTTACGGTTTCCGCGCCCCCG 26144
Qy        116 TAATCATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCA 175
Db      26143 TCATTAAAAAAATTAATCGGCAAAAACCATCAGGTGTATGCCTTTGTATCGGAGTTTTCCG 26084
Qy        176 CTGAAGATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACT 235
Db      26083 ACAATGAATTGGATATTATCAGGGAAATGGGGGTTACACCCGTTACCTACCGTTCAAACC 26024
Qy        236 CAAAGGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTA 295
Db      26023 GCAGCGGGCTGAACCCGTTTTCGGATATAAAATCCACCTTCCTCATCTTTAAAGAACTCA 25964
Qy        296 AGGATATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTA 355
Db      25963 AAAAAATATCGCCGGATTTGGTTTTCCCTTATTTCGCAAAACCCGTGATTTTCGGCACTT 25904
Qy        356 TTGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATG 415
Db      25903 TTGCCGCAAAACTGGCAGGCGTGCCCAGAATCGTCGGGATGCTGGAAGGTTTGGGATTCG 25844
Qy        416 CCTTCACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAA 475
Db      25843 CATTTACCCCGCAGCCGGAAGGCATACCGTTAAAAACAAAAATCATAAAGGGGATTTTGA 25784
Qy        476 TTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATG 535
Db      25783 TTGCCTTATACCGCATTGCCCTGCCGATGTTGGAAAGCCTGATTGTATTAAACCCCGACG 25724
Qy        536 ATAAAAAAGATTTAATCGATCAGTATAATATTAA---AGCTAAGGTAACAGTGTTAGGTG 592
Db      25723 ACAAAGACGAACTGACGGACAAATACGGCATCAAAATAAAAAACATCCATATTTTGGGCG 25664

Qy        593 GGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCG----------AAAG 640
Db      25663 GAATCGGTCTGGATTTGCGGCAATATCCTTATTCCGAGGCGGATATTCCCGATGAAAAAG 25604
Qy        641 AGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTA 700
Db      25603 AACCCGTAAAATTCCTCTTTATCGGCAGATTTCTGAAAGAAAAGGGGATTGATGATTTTA 25544
Qy        701 TTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAG 760
Db      25543 TTCGGGCGGCGGAACAGGTTAAGGACAAATACCCCGATACGGTTTTTACCGCTTTGGGCG 25484
Qy        761 GTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAG 820
Db      25483 CAATCGACA---------AATCACGCGGGGGGGGGGGCGATTTGGAACGGCTTGCCGCCC 25433
Qy        821 AACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTT 880
Db      25432 GCGATATTATCCGTTTCCCCGGTTTTGTGAACAATGTTTCCGAAGTGATAAAAGAACATC 25373
Qy        881 CTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTA 940
Db      25372 ATATATTCGTATTGCCGTCTTATTATAGGGAAGGCGTTCCCCGAAGCACTCAGGAGGCAA 25313
Qy        941 TGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATG 1000
Db      25312 TGGCCGTCGGCAGGGCAGTGATTACGACGGATGTCCCCGGATGCAGGGAAACGGTCGCCG 25253
Qy       1001 ATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGA 1060
Db      25252 ACAAGGTCAACGGCTTCCTGATCGAGCCTTGGAATCCCCGCATCTTGGCCGAAAAAATGA 25193
Qy       1061 AATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTG 1120
Db      25192 TTTATTTTATCGAAAACAGGGAAGCCGTCCGCCTGATGGGGAATGCAAGTTATGCGATTG 25133
Qy       1121 CAGAAAAAAACTTTGATGC 1139
Db      25132 CCAAAGATAAATTCGATGC 25114


RESULT 12
AX043922/c
LOCUS       AX043922               349980 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 1 from Patent WO0066791.
ACCESSION   AX043922
VERSION     AX043922.1  GI:11342850
KEYWORDS    .
SOURCE      Neisseria meningitidis
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Neisseria.
REFERENCE   1
  AUTHORS   Pizza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
            Masignani,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
            Scarlato,V., Rappuoli,R., Frazer,C.M. and Grandi,G.
  TITLE     Neisseria genomic sequences and methods of their use
  JOURNAL   Patent: WO 0066791-A 1 09-NOV-2000;
            CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Neisseria meningitidis"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:487"
                     /note="sequence too long, cut in 8 pieces.
                     seq 1: 1 to 349980 349980 bases
                     seq 108: 300001 to 649980 349980 bases
                     seq 109: 600001 to 949980 349980 bases
                     seq 110: 900001 to 1249980 349980 bases
                     seq 111: 1200001 to 1549980 349980 bases
                     seq 112: 1500001 to 1849980 349980 bases
                     seq 113: 1800001 to 2149980 349980 bases
                     seq 114: 2100001 to 2272325 172325 bases"
```

ORIGIN

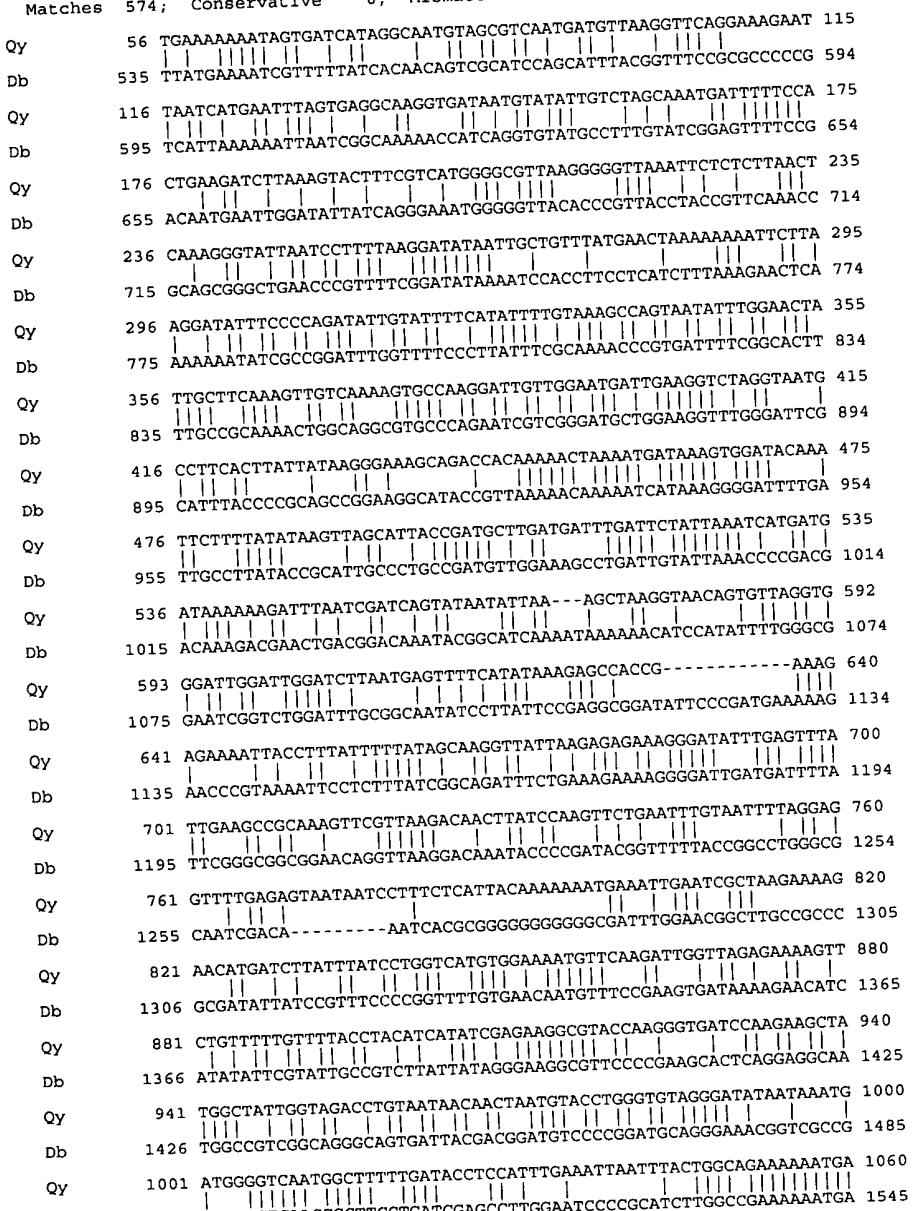
Query Match 18.6%; Score 222.6; DB 6; Length 349980;
Best Local Similarity 52.4%; Pred. No. 4e-25;
Matches 576; Conservative 0; Mismatches 499; Indels 24; Gaps 3;



RESULT 13
NMU73942
LOCUS NMU73942 2963 bp DNA linear BCT 14-SEP-1998
DEFINITION Neisseria meningitidis 3-oxoacyl-[acyl-carrier-protein] synthase II (fabF) gene, partial cds; and glycosyltransferase (pglA) gene, complete cds.
ACCESSION U73942
VERSION U73942.1 GI:3599930
KEYWORDS .
SOURCE Neisseria meningitidis
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 2963)
AUTHORS Jennings,M.P., Virji,M., Evans,D., Foster,V., Srikhanta,Y.N., Steeghs,L., van der Ley,P. and Moxon,E.R.
TITLE Identification of a novel gene involved in pilin glycosylation in Neisseria meningitidis
JOURNAL Mol. Microbiol. 29 (4), 975-984 (1998)
PUBMED 9767566
REFERENCE 2 (bases 1 to 2963)
AUTHORS Jennings,M.P.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1996) School of Biomolecular and Biomedical Science, Faculty of Science and Technology, Griffith University, Kessels Road, Nathan, Brisbane, Queensland 4111, Australia
FEATURES Location/Qualifiers
source 1. .2963
/organism="Neisseria meningitidis"
/mol_type="genomic DNA"
/strain="MC58"
/db_xref="taxon:487"
gene <1. .232
/gene="fabF"
CDS <1. .232
/gene="fabF"
/codon_start=2
/transl_table=11
/product="3-oxoacyl-[acyl-carrier-protein] synthase II"
/protein_id="AAC35425.1"
/db_xref="GI:3599931"
/translation="HLLGAAGGVEAVYSILAIHDGKIPPTINIFEQDVEAGCDLDYCA
NEARDAEIDVAISNSFGFGGTNGTLVFKRFKG"
gene 537. .1667
/gene="pglA"
CDS 537. .1667
/gene="pglA"
/function="involved in pilin glycosylation"
/note="similar to Shigella dysenteriae RfpB encoded by GenBank Accession Number M96064"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAC35426.1"
/db_xref="GI:3599932"
/translation="MKIVFITTVASSIYGFRAPVIKKLIGKNHQVYAFVSEFSDNELD
IIREMGVTPVTYRSNRSGLNPFSDIKSTFLIFKELKKISPDLVFPYFAKPVIFGTFAA
KLAGVPRIVGMLEGLGFAFTPQPEGIPLKTKIIKGILIALYRIALPMLESLIVLNPDD
KDELTDKYGIKIKNIHILGGIGLDLRQYPYSEADIPDEKEPVKFLFIGRFLKEKGIDD
FIRAAEQVKDKYPDTVFTGLGAIDKSRGGGGDLERLAARDIIRFPGFVNNVSEVIKEH
HIFVLPSYYREGVPRSTQEAMAVGRAVITTDVPGCRETVADKVNGFLIEPWNPRILAE
KMIYFIENREAVRLMGNASYAIAKDKFDAEKVDLKLLDILKA"

ORIGIN

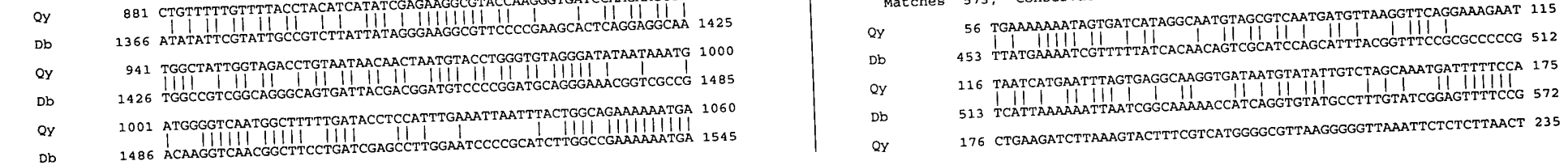
Query Match 18.4%; Score 219.4; DB 1; Length 2963;
Best Local Similarity 52.2%; Pred. No. 7.5e-24;
Matches 574; Conservative 0; Mismatches 501; Indels 24; Gaps 3;

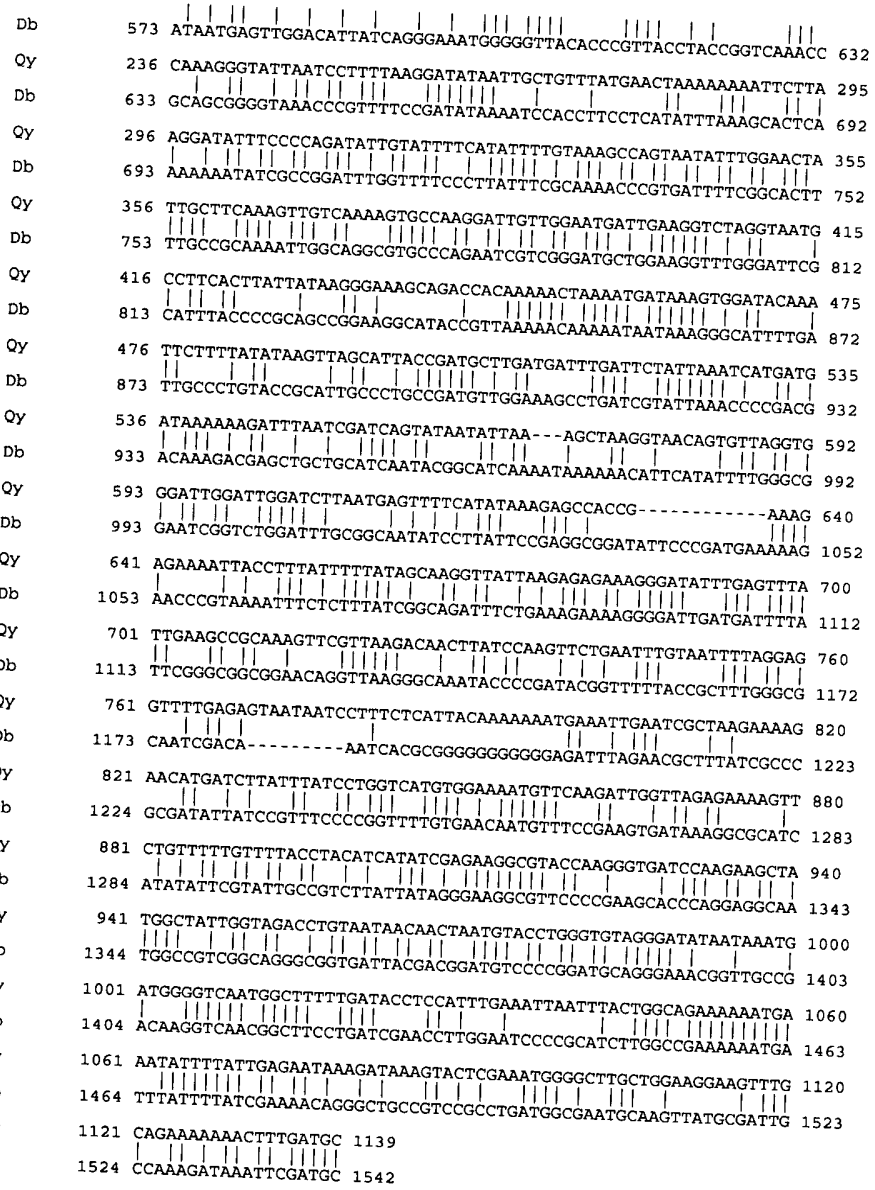


RESULT 14
AF485418
LOCUS AF485418 1585 bp DNA linear BCT 19-JUL-2002
DEFINITION Neisseria gonorrhoeae strain FA1090 pilin glycosyl transferase A (pgtA) gene, pgtA-2 allele, complete cds.
ACCESSION AF485418
VERSION AF485418.1 GI:20219021
KEYWORDS .
SOURCE Neisseria gonorrhoeae
ORGANISM Neisseria gonorrhoeae
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 1585)
AUTHORS Banerjee,A., Wang,R., Supernavage,S.L., Ghosh,S.K., Parker,J., Ganesh,N.F., Wang,P.G., Gulati,S. and Rice,P.A.
TITLE Implications of Phase Variation of a Gene (pgtA) Encoding a Pilin Galactosyl Transferase in Gonococcal Pathogenesis
JOURNAL J. Exp. Med. 196 (2), 147-162 (2002)
PUBMED 12119340
REFERENCE 2 (bases 1 to 1585)
AUTHORS Banerjee,A.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2002) Biology, Catholic Univ. of America, 620 Michigan Avenue, NE, Washington, DC 20064, USA
FEATURES Location/Qualifiers
source 1. .1585
/organism="Neisseria gonorrhoeae"
/mol_type="genomic DNA"
/strain="FA1090"
/db_xref="taxon:485"
gene 455. .1585
/gene="pgtA"
/allele="2"
CDS 455. .1585
/gene="pgtA"
/allele="2"
/note="galactosyl transferase"
/codon_start=1
/transl_table=11
/product="pilin glycosyl transferase A"
/protein_id="AAM15778.1"
/db_xref="GI:20219022"
/translation="MKIVFITTVASSIYGFRAPVIKKLIGKNHQVYAFVSEFSDNELD IIREMGVTPVTYRSNRSGVNPFSDIKSTFLIFKALKKISPDLVFPYFAKPVIFGTFAA KLAGVPRIVGMLEGLGFAFTPQPEGIPLKTKIIKGILIALYRIALPMLESLIVLNPDD KDELLHQYGIKIKNIHILGGIGLDLRQYPYSEADIPDEKEPVKFLFIGRFLKEKGIDD FIRAAEQVKGKYPDTVFTALGAIDKSRGGGGDLERFIARDIIRFPGFVNNVSEVIKAH HIFVLPSYYREGVPRSTQEAMAVGRAVITTDVPGCRETVADKVNGFLIEPWNPRILAE KMIYFIENRAAVRLMANASYAIAKDKFDAEKVDLKFLDILKA"

ORIGIN

Query Match 18.2%; Score 217.8; DB 1; Length 1585;
Best Local Similarity 52.1%; Pred. No. 1.7e-23;
Matches 573; Conservative 0; Mismatches 502; Indels 24; Gaps 3;



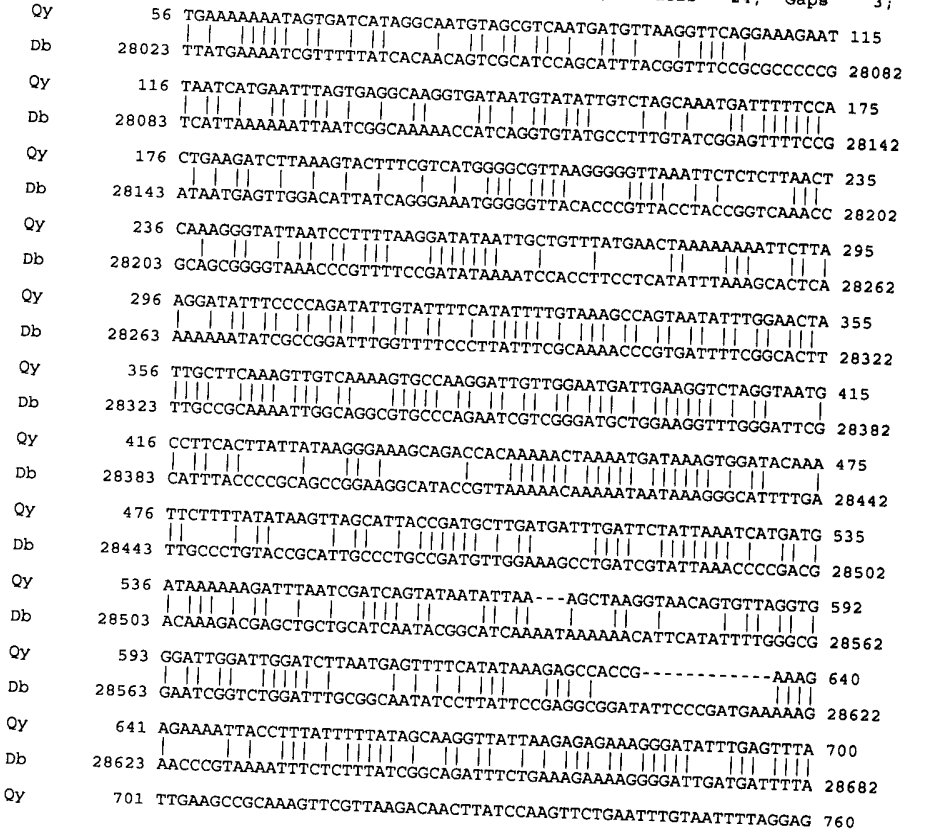


RESULT 15
AE004969_17
WPCOMMENT
Sequence split into 22 fragments LOCUS AE004969 Accession AE004969

Fragment Name	Begin	End
AE004969_00	1	110000
AE004969_01	100001	210000
AE004969_02	200001	310000
AE004969_03	300001	410000
AE004969_04	400001	510000
AE004969_05	500001	610000
AE004969_06	600001	710000
AE004969_07	700001	810000
AE004969_08	800001	910000
AE004969_09	900001	1010000
AE004969_10	1000001	1110000
AE004969_11	1100001	1210000
AE004969_12	1200001	1310000
AE004969_13	1300001	1410000
AE004969_14	1400001	1510000
AE004969_15	1500001	1610000
AE004969_16	1600001	1710000
AE004969_17	1700001	1810000
AE004969_18	1800001	1910000
AE004969_19	1900001	2010000
AE004969_20	2000001	2110000
AE004969_21	2100001	2153922

Continuation (18 of 22) of AE004969 from base 1700001 (AE004969 Neisseria gonorrhoeae FA

Query Match 18.2%; Score 217.8; DB 1; Length 110000;
Best Local Similarity 52.1%; Pred. No. 3.3e-24;
Matches 573; Conservative 0; Mismatches 502; Indels 24; Gaps 3;



```
Db      28683 TTCGGGCGGCGGAACAGGTTAAGGGCAAATACCCCGATACGGTTTTTACCGCTTTGGGCG 28742

Qy        761 GTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAG 820

Db      28743 CAATCGACA---------AATCACGCGGGGGGGGGGGAGATTTAGAACGCTTTATCGCCC 28793

Qy        821 AACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTT 880

Db      28794 GCGATATTATCCGTTTCCCCGGTTTTGTGAACAATGTTTCCGAAGTGATAAAGGCGCATC 28853

Qy        881 CTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTA 940

Db      28854 ATATATTCGTATTGCCGTCTTATTATAGGGAAGGCGTTCCCCGAAGCACCCAGGAGGCAA 28913

Qy        941 TGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATG 1000

Db      28914 TGGCCGTCGGCAGGGCGGTGATTACGACGGATGTCCCCGGATGCAGGGAAACGGTTGCCG 28973

Qy       1001 ATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGA 1060

Db      28974 ACAAGGTCAACGGCTTCCTGATCGAACCTTGGAATCCCCGCATCTTGGCCGAAAAAATGA 29033

Qy       1061 AATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTG 1120

Db      29034 TTTATTTTATCGAAAACAGGGCTGCCGTCCGCCTGATGGCGAATGCAAGTTATGCGATTG 29093

Qy       1121 CAGAAAAAAACTTTGATGC 1139

Db      29094 CCAAAGATAAATTCGATGC 29112
```

Search completed: May 10, 2006, 20:34:10
Job time : 4340.97 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:42:47 ; Search time 557.376 Seconds
(without alignments)
14276.972 Million cell updates/sec

Title: US-09-423-093B-1_COPY_739_1932
Perfect score: 1194
Sequence: 1 atgttactatgctgcataca.....aatcaaataatgatttttga 1194

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1194	100.0	14516	2	AAX06748	Aax06748 E. coli O
	2	1194	100.0	14516	3	AAZ56375	Aaz56375 Escherich
	3	222.6	18.6	52253	3	AAA81478	Aaa81478 N. mening
c	4	222.6	18.6	110000	3	AAA81490_02	Continuation (3 of
c	5	222.6	18.6	349980	3	AAF21544	Aaf21544 Neisseria
	6	217.8	18.2	1227	10	ABZ39815	Abz39815 N. gonorr
	7	217.8	18.2	1227	10	ABZ41249	Abz41249 N. gonorr
	8	170	14.2	1173	9	ADA29503	Ada29503 DNA encod
	9	116.4	9.7	17433	13	ADW38769	Adw38769 Shigella
	10	107.2	9.0	14139	14	AEA81303	Aea81303 E. coli O
	11	105	8.8	13234	14	AEA29363	Aea29363 Shigella
	12	103.4	8.7	11717	14	AEC31411	Aec31411 E. coli t
	13	88.4	7.4	17623	12	ADF77651	Adf77651 Francisel
	14	88.4	7.4	22401	12	ADF77691	Adf77691 Francisel
	15	78.4	6.6	18900	13	ADW00484	Adw00484 Escherich
	16	73	6.1	1134	8	ACA50987	Aca50987 Prokaryot
	17	72	6.0	13343	14	AEC21874	Aec21874 E. coli O
	18	70.2	5.9	1092	8	ACA28175	Aca28175 Prokaryot
	19	69	5.8	1143	11	ACH95187	Ach95187 Klebsiell
	20	67.8	5.7	1080	13	ADS45685	Ads45685 Bacterial
	21	67.8	5.7	13402	14	AEA37602	Aea37602 Shigella
	22	66.8	5.6	12110	14	AEA33399	Aea33399 S. boydii
	23	65.6	5.5	1089	8	ACA29324	Aca29324 Prokaryot
	24	65.4	5.5	110000	2	AAV21209_03	Continuation (4 of
	25	63	5.3	1122	10	ACF70466	Acf70466 Photorhab
	26	63	5.3	69727	10	ACF65374	Acf65374 Photorhab
	27	63	5.3	110000	10	ACF67367_35	Continuation (36 o
c	28	62	5.2	349980	5	AAH41226	Aah41226 Pyrococcu
	29	59.8	5.0	13074	14	AEC21818	Aec21818 E. coli O
	30	59.8	5.0	16217	6	ABL32624	Abl32624 Human imm
	31	58.6	4.9	8056	8	ABZ10246	Abz10246 Haematopo
	32	58.4	4.9	1134	10	ADF03016	Adf03016 Bacterial
c	33	57.8	4.8	8056	8	ABZ10246	Abz10246 Haematopo
	34	57.6	4.8	5999	6	ABK39960	Abk39960 Human che
	35	57.6	4.8	5999	6	ABL32558	Abl32558 Human imm
	36	57.4	4.8	15373	6	ABL32466	Abl32466 Human imm
	37	56.8	4.8	115218	8	ACA64845	Aca64845 Human HNR
	38	56.6	4.7	7458	3	AAA70106	Aaa70106 Plasmodiu
c	39	56.2	4.7	2000	11	ACL37108	Acl37108 Rice stre
	40	55.4	4.6	8056	8	ABZ10100	Abz10100 Haematopo
c	41	54.8	4.6	2000	11	ACL35887	Acl35887 Rice stre
c	42	54	4.5	50000	6	ABL56201	Abl56201 AmEPV gen
	43	53.8	4.5	1191	10	ADC91458	Adc91458 E. faeciu
c	44	53.8	4.5	1683	3	AAC79598	Aac79598 Virulence
c	45	53.8	4.5	1683	6	ABQ83475	Abq83475 Pasteurel

ALIGNMENTS

RESULT 1
AAX06748
ID AAX06748 standard; DNA; 14516 BP.
XX
AC AAX06748;
XX
DT 26-APR-1999 (first entry)
XX
DE E. coli O111 antigen gene cluster.
XX
KW O antigen; O111 antigen; diagnosis; wbdH gene; wzx gene; wzy gene;
KW wbdM gene; rfb gene; transferase; polymerase; flippase; diarrhoea;
KW haemorrhagic colitis; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 739. .1932
FT /*tag= a
FT /product= "galactosyl_transferase"
FT /note= "wbdH gene (ORF1), this region is specifically
FT claimed in Claim 6, and encodes AAW88300"
FT CDS 1983. .3101
FT /*tag= b
FT /note= "gmd gene (ORF2); encodes AAW88301"
FT CDS 3174. .3623
FT /*tag= c
FT /note= "wbdI gene (ORF3); encodes AAW88302"
FT CDS 3657. .5105
FT /*tag= d
FT /note= "manC gene (ORF4); encodes AAW88303"
FT CDS 5114. .6484
FT /*tag= e
FT /note= "manB gene (ORF5); encodes AAW88304"
FT CDS 6521. .7444
FT /*tag= f
FT /note= "wbdJ gene (ORF6); encodes AAW88305"
FT CDS 7441. .8607
FT /*tag= g
FT /note= "wbdK gene (ORF7); encodes AAW88306"
FT CDS 8646. .9911
FT /*tag= h

```
FT                   /product= "O_antigen_flippase"
FT                   /note= "wzx gene (ORF8), this region is specifically
FT                   claimed in Claim 6; encodes AAW88307"
FT   CDS             9901. .10953
FT                   /*tag= i
FT                   /product= "O_antigen_polymerase"
FT                   /note= "wzy gene (ORF9), this region is specifically
FT                   claimed in Claim 6; encodes AAW88308"
FT   CDS             10931. .11824
FT                   /*tag= j
FT                   /product= "colitose_or_glucose_transferase"
FT                   /note= "wbdL gene (ORF10); encodes AAW88309"
FT   CDS             11821. .12945
FT                   /*tag= k
FT                   /product= "colitose or glucose transferase"
FT                   /note= "wbdM gene (ORF11), this region is specifically
FT                   claimed in Claim 6; encodes AAW88310"
FT   CDS             13124. .14516
FT                   /*tag= l
FT                   /note= "gnd gene (partial sequence); encodes AAW88311"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;    98WO-AU000315.
XX
PR   01-MAY-1997;    97AU-00006545.
PR   22-JUL-1997;    97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR, Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88300, AAW88301, AAW88302, AAW88303, AAW88304, AAW88305,
DR   AAW88306, AAW88307, AAW88308, AAW88309, AAW88310, AAW88311.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 6; Fig 7; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Escherichia coli O111 O antigen. The gene cluster was
CC   isolated from a cosmid gene bank of E. coli M92 chromosomal DNA by
CC   screening with E. coli O111 antiserum. Open reading frames within the
CC   gene cluster were identified on the basis of homology to known sequences.
CC   The genes encode proteins (see AAW88300-11) involved the synthesis of
CC   sugars present in the polysaccharide antigen, and in the transport or
CC   processing of polysaccharide or oligosaccharide units. The use of nucleic
CC   acid molecules derived from particular assembly and transport genes,
CC   particularly wbd (transferase), wzx (flippase) and wzy (polymerase)
CC   genes, within O antigen gene clusters improves the specificity of methods
CC   for the detection and identification of O antigens, e.g. in testing food-
CC   or faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of enteropathogenic E. coli strains that cause
CC   diarrhoea and haemorrhagic colitis
XX
SQ   Sequence 14516 BP; 4582 A; 2221 C; 2971 G; 4742 T; 0 U; 0 Other;
```

Query Match 100.0%; Score 1194; DB 2; Length 14516;
Best Local Similarity 100.0%; Pred. No. 2e-230;
Matches 1194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 60
Db     739 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 798

Qy      61 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 120
Db     799 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 858

Qy     121 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 180
Db     859 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 918

Qy     181 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAG 240
Db     919 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAG 978

Qy     241 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 300
Db     979 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 1038

Qy     301 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 360
Db    1039 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 1098

Qy     361 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 420
Db    1099 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 1158

Qy     421 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 480
Db    1159 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 1218

Qy     481 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 540
Db    1219 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 1278

Qy     541 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 600
Db    1279 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 1338

Qy     601 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 660
Db    1339 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 1398

Qy     661 ATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTT 720
Db    1399 ATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTT 1458

Qy     721 AAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCT 780
Db    1459 AAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCT 1518

Qy     781 TTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCT 840
Db    1519 TTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCT 1578

Qy     841 GGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACA 900
Db    1579 GGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACA 1638

Qy     901 TCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTA 960
Db    1639 TCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTA 1698

Qy     961 ATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTG 1020
Db    1699 ATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTG 1758

Qy    1021 ATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAA 1080
Db    1759 ATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAA 1818

Qy    1081 GATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCT 1140
Db    1819 GATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCT 1878

Qy    1141 TTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1194
Db    1879 TTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1932
```

```
RESULT 2
AAZ56375
ID   AAZ56375 standard; DNA; 14516 BP.
XX
AC   AAZ56375;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:45.
XX
KW   Flagellin; fliC; antigen; detection; ds.
XX
OS   Escherichia coli.
XX
PN   WO9961458-A1.
XX
PD   02-DEC-1999.
XX
PF   21-MAY-1999;   99WO-AU000385.
XX
PR   21-MAY-1998;   98AU-00003634.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 2000-072598/06.
XX
PT   Novel nucleic acid molecule useful for the detection of flagellated
PT   bacterial strains in food, feces, etc.
XX
PS   Claim 3; Page 216-221; 245pp; English.
XX
CC   AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all or
CC   part of an Escherichia coli flagellin protein except a protein expressed
CC   by E. coli H1, H7, H12 or H48 type strains. The present invention also
CC   describes a method of detecting the presence of E. coli of a particular H
CC   serotype in a sample, comprising specifically hybridising a nucleic acid,
CC   preferably at least a pair, derived from a flagellating gene, specific
CC   for a particular flagellin gene associated with the H serotype, to any
CC   E.coli in the sample which contain the gene, and detecting any hybridised
CC   molecules, identifying the presence of that serotype in the sample. (I)
CC   are useful for: (1) detecting the presence of E. coli of H serotype in a
CC   sample by hybridising at least one or a pair of (I) to any E. coli in the
CC   sample and detecting the hybridised nucleic acid molecules; and (2) for
CC   detecting the presence of both O and H-serotypes of E. coli by
CC   hybridising at least one or a pair of (I) to any E. coli present in the
CC   sample and detecting the hybridised nucleic acid molecules. (I) is
CC   particularly useful for detecting the combination of O and H antigen.
CC   Hybridised (I) when using at least one (I) is detected by southern blot
CC   analysis and, when using a pair of (I), is detected by polymerase chain
CC   reaction (PCR). AAZ56399 to AAZ56420 represent primers used in the
CC   exemplification of the present invention
XX
SQ   Sequence 14516 BP; 4582 A; 2221 C; 2971 G; 4742 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1194;  DB 3;  Length 14516;
  Best Local Similarity   100.0%;  Pred. No. 2e-230;
  Matches 1194;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 798

Qy         61 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 858

Qy        121 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        859 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 918

Qy        181 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        919 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAG 978

Qy        241 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        979 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 1038

Qy        301 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1039 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 1098

Qy        361 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1099 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 1158

Qy        421 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1159 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 1218

Qy        481 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1219 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 1278

Qy        541 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1279 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 1338

Qy        601 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1339 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 1398

Qy        661 ATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1399 ATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTT 1458

Qy        721 AAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1459 AAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCT 1518

Qy        781 TTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1519 TTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCT 1578

Qy        841 GGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1579 GGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACA 1638

Qy        901 TCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1639 TCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTA 1698

Qy        961 ATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1699 ATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTG 1758

Qy       1021 ATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1759 ATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAA 1818

Qy       1081 GATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1819 GATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCT 1878

Qy       1141 TTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1879 TTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1932


RESULT 3
AAA81478
ID   AAA81478 standard; DNA; 52253 BP.
XX
```

AC AAA81478;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_26 SEQ ID NO:26.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US023573.
XX
PR 09-OCT-1998; 98US-0103794P.
PR 30-APR-1999; 99US-0132068P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be used
PT in the diagnosis and treatment of N. meningitidis infection and other
PT Neisserial infections, for example, N.gonorrhoea.
XX
PS Claim 7; Page 532-547; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic proteins
CC from Neisseria genomic sequences. AAA81453 to AAA82414 represent
CC specifically claimed Neisseria meningitidis genomic DNA sequences;
CC AAA81260 to AAA81303 and AAB25620 to AAB25663 represent Neisseria DNA
CC sequences and their corresponding proteins; AAA81254 to AAA81259 and
CC AAA81304 to AAA81321 represent PCR primers used in the isolation of
CC Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent
CC Neisseria meningitidis MenB polynucleotide ORF sequences, which are all
CC used in the exemplification of the present invention. The nucleic acid
CC sequences, protein sequences, and antibodies against them, can be used in
CC the manufacture of a composition. The composition can be used as a
CC medicament (or in the manufacture of a medicament) for treating,
CC preventing or diagnosing infection due to Neisserial bacteria. For
CC example, some of the identified proteins could be components of vaccines
CC against Meningococcus B; against all serotypes; and/or against all
CC pathogenic Neissariae. Identification of sequences from the bacterium
CC will also facilitate production of biological probes, particularly
CC organism-specific probes. Attempts to make efficacious Meningococcus B
CC vaccines have failed mainly due to antigen tolerance. Multivalent
CC vaccines have also been tried but none have successfully overcome
CC antigenic variability. The provision of further, complete sequences may
CC provide an opportunity to identify secreted or surface exposed proteins
CC that may be presumed targets for the immune system and which are not
CC antigenically variable or at least more conserved than other more
CC variable regions
XX
SQ Sequence 52253 BP; 12877 A; 15510 C; 13103 G; 10761 T; 0 U; 2 Other;

Query Match 18.6%; Score 222.6; DB 3; Length 52253;
Best Local Similarity 52.4%; Pred. No. 4.8e-35;
Matches 576; Conservative 0; Mismatches 499; Indels 24; Gaps 3;

```
Qy        56 TGAAAAAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAAT 115
Db      1908 TTATGAAAATCGTTTTTATCACAACAGTCGCATCCAGCATTTACGGTTTCCGCGCCCCCG 1967

Qy       116 TAATCATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCA 175
```

Db 1968 TCATTAAAAAATTAATCGGCAAAAACCATCAGGTGTATGCCTTTGTATCGGAGTTTTCCG 2027

```
Qy       176 CTGAAGATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACT 235
Db      2028 ACAATGAATTGGATATTATCAGGGAAATGGGGGTTACACCCGTTACCTACCGTTCAAACC 2087

Qy       236 CAAAGGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTA 295
Db      2088 GCAGCGGGCTGAACCCGTTTTCGGATATAAAATCCACCTTCCTCATCTTTAAAGAACTCA 2147

Qy       296 AGGATATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTA 355
Db      2148 AAAAAATATCGCCGGATTTGGTTTTCCCTTATTTCGCAAAACCCGTGATTTTCGGCACTT 2207

Qy       356 TTGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATG 415
Db      2208 TTGCCGCAAAACTGGCAGGCGTGCCCAGAATCGTCGGGATGCTGGAAGGTTTGGGATTCG 2267

Qy       416 CCTTCACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAA 475
Db      2268 CATTTACCCCGCAGCCGGAAGGCATACCGTTAAAAACAAAAATCATAAAGGGGATTTTGA 2327

Qy       476 TTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATG 535
Db      2328 TTGCCTTATACCGCATTGCCCTGCCGATGTTGGAAAGCCTGATTGTATTAAACCCCGACG 2387

Qy       536 ATAAAAAAGATTTAATCGATCAGTATAATATTAA---AGCTAAGGTAACAGTGTTAGGTG 592
Db      2388 ACAAAGACGAACTGACGGACAAATACGGCATCAAAATAAAAAACATCCATATTTTGGGCG 2447

Qy       593 GGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCG----------AAAG 640
Db      2448 GAATCGGTCTGGATTTGCGGCAATATCCTTATTCCGAGGCGGATATTCCCGATGAAAAAG 2507

Qy       641 AGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTA 700
Db      2508 AACCCGTAAAATTCCTCTTTATCGGCAGATTTCTGAAAGAAAAGGGGATTGATGATTTTA 2567

Qy       701 TTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAG 760
Db      2568 TTCGGGCGGCGGAACAGGTTAAGGACAAATACCCCGATACGGTTTTTACCGCTTTGGGCG 2627

Qy       761 GTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAG 820
Db      2628 CAATCGACA---------AATCACGCGGGGGGGGGGGCGATTTGGAACGGCTTGCCGCCC 2678

Qy       821 AACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTT 880
Db      2679 GCGATATTATCCGTTTCCCCGGTTTTGTGAACAATGTTTCCGAAGTGATAAAAGAACATC 2738

Qy       881 CTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTA 940
Db      2739 ATATATTCGTATTGCCGTCTTATTATAGGGAAGGCGTTCCCCGAAGCACTCAGGAGGCAA 2798

Qy       941 TGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATG 1000
Db      2799 TGGCCGTCGGCAGGGCAGTGATTACGACGGATGTCCCCGGATGCAGGGAAACGGTCGCCG 2858

Qy      1001 ATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGA 1060
Db      2859 ACAAGGTCAACGGCTTCCTGATCGAGCCTTGGAATCCCCGCATCTTGGCCGAAAAAATGA 2918

Qy      1061 AATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTG 1120
Db      2919 TTTATTTTATCGAAAACAGGGAAGCCGTCCGCCTGATGGGGAATGCAAGTTATGCGATTG 2978

Qy      1121 CAGAAAAAAACTTTGATGC 1139
Db      2979 CCAAAGATAAATTCGATGC 2997

RESULT 4
AAA81490_02/c
```

```
Continuation (3 of 15) of AAA81490 from base 200001 (N. meningitidis B full length genom
WP Sequence split into 15 fragments  LOCUS AAA81490 Accession Aaa81490
WP      Fragment Name       Begin      End
WP      AAA81490_00             1    110000
WP      AAA81490_01        100001    210000
WP      AAA81490_02        200001    310000
WP      AAA81490_03        300001    410000
WP      AAA81490_04        400001    510000
WP      AAA81490_05        500001    610000
WP      AAA81490_06        600001    710000
WP      AAA81490_07        700001    810000
WP      AAA81490_08        800001    910000
WP      AAA81490_09        900001   1010000
WP      AAA81490_10       1000001   1110000
WP      AAA81490_11       1100001   1210000
WP      AAA81490_12       1200001   1310000
WP      AAA81490_13       1300001   1410000
WP      AAA81490_14       1400001   1437668

  Query Match             18.6%;  Score 222.6;  DB 3;  Length 110000;
  Best Local Similarity   52.4%;  Pred. No. 5.3e-35;
  Matches  576;  Conservative    0;  Mismatches  499;  Indels   24;  Gaps    3;

Qy         56 TGAAAAAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAAT 115
Db      26195 TTATGAAAATCGTTTTTATCACAACAGTCGCATCCAGCATTTACGGTTTCCGCGCCCCCG 26136

Qy        116 TAATCATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCA 175
Db      26135 TCATTAAAAAATTAATCGGCAAAAACCATCAGGTGTATGCCTTTGTATCGGAGTTTTCCG 26076

Qy        176 CTGAAGATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACT 235
Db      26075 ACAATGAATTGGATATTATCAGGGAAATGGGGGTTACACCCGTTACCTACCGTTCAAACC 26016

Qy        236 CAAAGGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTA 295
Db      26015 GCAGCGGGCTGAACCCGTTTTCGGATATAAAATCCACCTTCCTCATCTTTAAAGAACTCA 25956

Qy        296 AGGATATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTA 355
Db      25955 AAAAAATATCGCCGGATTTGGTTTTCCCTTATTTCGCAAAACCCGTGATTTTCGGCACTT 25896

Qy        356 TTGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATG 415
Db      25895 TTGCCGCAAAACTGGCAGGCGTGCCCAGAATCGTCGGGATGCTGGAAGGTTTGGGATTCG 25836

Qy        416 CCTTCACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAA 475
Db      25835 CATTTACCCCGCAGCCGGAAGGCATACCGTTAAAAACAAAAATCATAAAGGGGATTTTGA 25776

Qy        476 TTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATG 535
Db      25775 TTGCCTTATACCGCATTGCCCTGCCGATGTTGGAAAGCCTGATTGTATTAAACCCCGACG 25716

Qy        536 ATAAAAAAGATTTAATCGATCAGTATAATATTAA---AGCTAAGGTAACAGTGTTAGGTG 592
Db      25715 ACAAAGACGAACTGACGGACAAATACGGCATCAAAATAAAAAACATCCATATTTTGGGCG 25656

Qy        593 GGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCG----------AAAG 640
Db      25655 GAATCGGTCTGGATTTGCGGCAATATCCTTATTCCGAGGCGGATATTCCCGATGAAAAAG 25596

Qy        641 AGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTA 700
Db      25595 AACCCGTAAAATTCCTCTTTATCGGCAGATTTCTGAAAGAAAAGGGGATTGATGATTTTA 25536

Qy        701 TTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAG 760
Db      25535 TTCGGGCGGCGGAACAGGTTAAGGACAAATACCCCGATACGGTTTTTACCGCTTTGGGCG 25476

Qy        761 GTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAG 820
Db      25475 CAATCGACA---------AATCACGCGGGGGGGGGGGCGATTTGGAACGGCTTGCCGCCC 25425

Qy        821 AACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTT 880
Db      25424 GCGATATTATCCGTTTCCCCGGTTTTGTGAACAATGTTTCCGAAGTGATAAAAGAACATC 25365

Qy        881 CTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTA 940
Db      25364 ATATATTCGTATTGCCGTCTTATTATAGGGAAGGCGTTCCCCGAAGCACTCAGGAGGCAA 25305

Qy        941 TGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATG 1000
Db      25304 TGGCCGTCGGCAGGGCAGTGATTACGACGGATGTCCCCGGATGCAGGGAAACGGTCGCCG 25245

Qy       1001 ATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGA 1060
Db      25244 ACAAGGTCAACGGCTTCCTGATCGAGCCTTGGAATCCCCGCATCTTGGCCGAAAAAATGA 25185

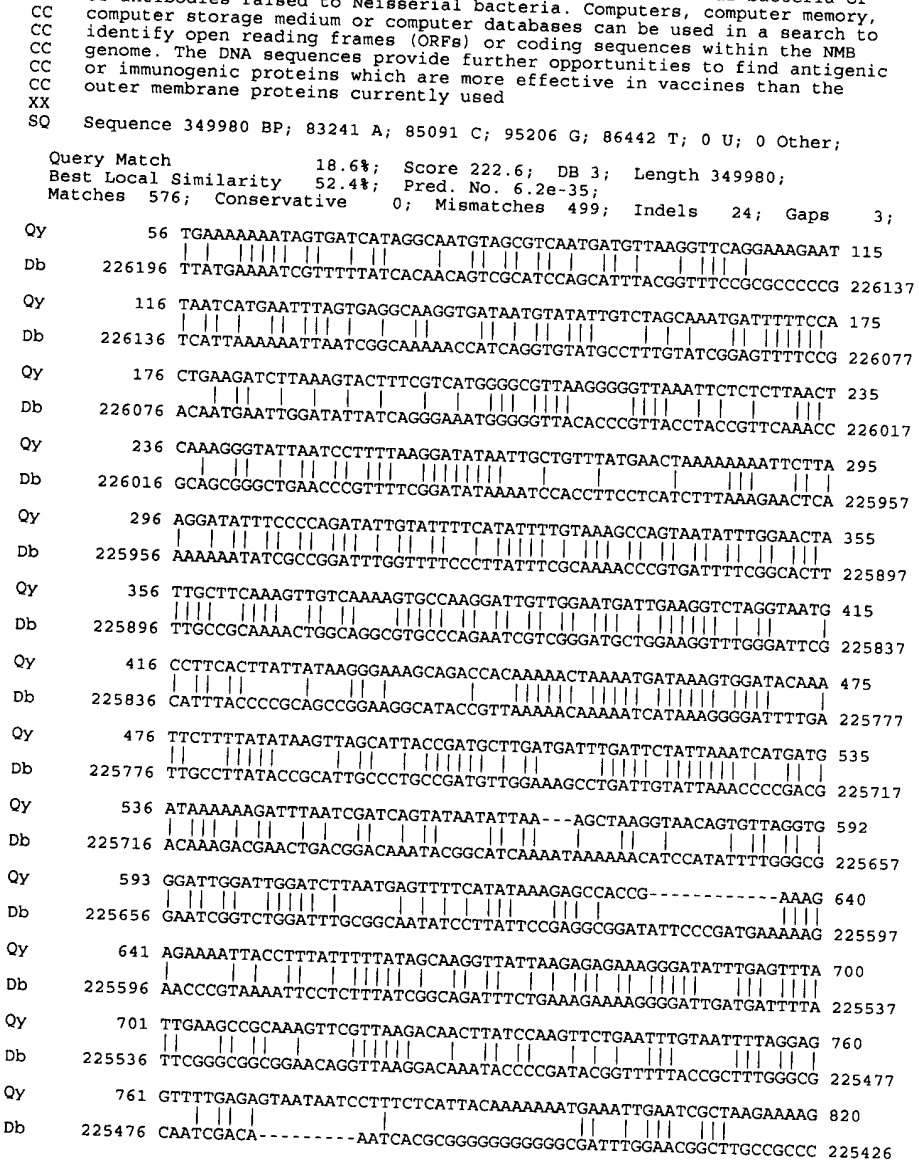
Qy       1061 AATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTG 1120
Db      25184 TTTATTTTATCGAAAACAGGGAAGCCGTCCGCCTGATGGGGAATGCAAGTTATGCGATTG 25125

Qy       1121 CAGAAAAAAACTTTGATGC 1139
Db      25124 CCAAAGATAAATTCGATGC 25106
```

```
RESULT 5
AAF21544/c
ID   AAF21544 standard; DNA; 349980 BP.
XX
AC   AAF21544;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Neisseria meningitidis B nucleotide sequence SEQ ID NO:1.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW   diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW   ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200066791-A1.
XX
PD   09-NOV-2000.
XX
PF   08-MAR-2000; 2000WO-US005928.
XX
PR   30-APR-1999;   99US-0132068P.
PR   08-OCT-1999;   99WO-US023573.
PR   28-FEB-2000; 2000GB-00004695.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Pizza M,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Frazer CM,  Grandi G;
XX
DR   WPI; 2000-647603/62.
XX
PT   Neisseria meningitidis B full length genome sequence and open reading
PT   frames are used to detect, treat and prevent Neisserial infections.
XX
PS   Claim 7; Appendix A; 692pp; English.
XX
CC   The present invention describes the full length genome of Neisseria
CC   meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613
CC   represent fragments of the NMB genomic sequence, as the sequence was too
CC   long to go in a record on its own it was split into 8 sequences which
CC   overlap each other at the beginning and end of each sequence by 49980 bp
CC   (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of
```

CC AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of
CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins
CC given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR
CC primers which are used in the exemplification of the present invention.
CC The NMB genome and fragments from it have antibacterial activity, and can
CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins
CC and/or antibodies which binds to the proteins can be used in compositions
CC for treating or preventing infection due to Neisserial bacteria or as a
CC diagnostic reagent for detecting the presence of Neisserial bacteria or
CC of antibodies raised to Neisserial bacteria. Computers, computer memory,
CC computer storage medium or computer databases can be used in a search to
CC identify open reading frames (ORFs) or coding sequences within the NMB
CC genome. The DNA sequences provide further opportunities to find antigenic
CC or immunogenic proteins which are more effective in vaccines than the
CC outer membrane proteins currently used
XX
SQ Sequence 349980 BP; 83241 A; 85091 C; 95206 G; 86442 T; 0 U; 0 Other;

Query Match 18.6%; Score 222.6; DB 3; Length 349980;
Best Local Similarity 52.4%; Pred. No. 6.2e-35;
Matches 576; Conservative 0; Mismatches 499; Indels 24; Gaps 3;





RESULT 6
ABZ39815
ID ABZ39815 standard; DNA; 1227 BP.
XX
AC ABZ39815;
XX
DT 07-MAR-2003 (first entry)
XX
DE N. gonorrhoeae nucleotide sequence SEQ ID 4219.
XX
KW Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX
OS Neisseria gonorrhoeae.
XX
PN WO200279243-A2.
XX
PD 10-OCT-2002.
XX
PF 12-FEB-2002; 2002WO-IB002069.
XX
PR 12-FEB-2001; 2001GB-00003424.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Fontana MR, Pizza M, Masignani V, Monaci E;
XX
DR WPI; 2003-058415/05.
DR P-PSDB; ABP78845.
XX
PT New protein from Neisseria gonorrheae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrheae infection.
XX
PS Disclosure; Page 495-496; 815pp; English.
XX
CC The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
CC molecules of the invention
XX
SQ Sequence 1227 BP; 328 A; 288 C; 296 G; 315 T; 0 U; 0 Other;

Query Match 18.2%; Score 217.8; DB 10; Length 1227;
Best Local Similarity 52.1%; Pred. No. 2.7e-34;

```
  Matches  573;  Conservative    0;  Mismatches  502;  Indels   24;  Gaps    3;

Qy          56 TGAAAAAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAAT 115
               | |  ||||| ||  | ||    |  || || || |  || |   | ||| |
Db          98 TTATGAAAATCGTTTTTATCACAACAGTCGCATCCAGCATTTACGGTTTCCGCGCCCCCG 157

Qy         116 TAATCATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCA 175
               | || |  || ||| | |  ||    || | || |||    | | |   || ||||||
Db         158 TCATTAAAAAATTAATCGGCAAAAACCATCAGGTGTATGCCTTTGTATCGGAGTTTTCCG 217

Qy         176 CTGAAGATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACT 235
                 | | ||  |  | |  |   | |  ||| ||||     ||||  | |      |||
Db         218 ATAATGAGTTGGACATTATCAGGGAAATGGGGGTTACACCCGTTACCTACCGGTCAAACC 277

Qy         236 CAAAGGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTA 295
                 |  ||  | || || |||   |||||||   |     |       ||   |||   || |
Db         278 GCAGCGGGGTAAACCCGTTTTCCGATATAAAATCCACCTTCCTCATATTTAAAGCACTCA 337

Qy         296 AGGATATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTA 355
               |  | || || || ||| | || ||  | |||||  | ||| || || || || || |||
Db         338 AAAAAATATCGCCGGATTTGGTTTTCCCTTATTTCGCAAAACCCGTGATTTTCGGCACTT 397

Qy         356 TTGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATG 415
               ||||  |||| ||| ||   |||||  || || || || ||| | |||||| | ||    |
Db         398 TTGCCGCAAAATTGGCAGGCGTGCCCAGAATCGTCGGGATGCTGGAAGGTTTGGGATTCG 457

Qy         416 CCTTCACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAA 475
               | || ||      |   || |       |   |||||| ||||| |||||| | ||    |
Db         458 CATTTACCCCGCAGCCGGAAGGCATACCGTTAAAAACAAAAATAATAAAGGGCATTTTGA 517

Qy         476 TTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATG 535
               ||    | ||     | ||  | |||||| | ||     ||||  |||||| |  || |
Db         518 TTGCCCTGTACCGCATTGCCCTGCCGATGTTGGAAAGCCTGATCGTATTAAACCCCGACG 577

Qy         536 ATAAAAAAGATTTAATCGATCAGTATAATATTAA---AGCTAAGGTAACAGTGTTAGGTG 592
               | ||| | ||  |  |  |||| ||   || ||   |   ||  |     | || || |
Db         578 ACAAAGACGAGCTGCTGCATCAATACGGCATCAAAATAAAAAACATTCATATTTTGGGCG 637

Qy         593 GGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCG----------AAAG 640
               | ||  ||  ||||| |     | | | | |||   ||| |                ||||
Db         638 GAATCGGTCTGGATTTGCGGCAATATCCTTATTCCGAGGCGGATATTCCCGATGAAAAAG 697

Qy         641 AGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTA 700
               |     | |  ||| | ||||| |  || ||  | | ||| || |||||   ||| ||||
Db         698 AACCCGTAAAATTTCTCTTTATCGGCAGATTTCTGAAAGAAAAGGGGATTGATGATTTTA 757

Qy         701 TTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAG 760
               ||   || ||  |    ||||||   |  || ||   | | |  |||     ||| || |
Db         758 TTCGGGCGGCGGAACAGGTTAAGGGCAAATACCCCGATACGGTTTTTACCGCTTTGGGCG 817

Qy         761 GTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAG 820
                 | || |          |                  ||  | |||    | |
Db         818 CAATCGACA---------AATCACGCGGGGGGGGGGGAGATTTAGAACGCTTTATCGCCC 868

Qy         821 AACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTT 880
                 ||  | |   ||  || ||| |||| | ||||||   ||   | || ||        |
Db         869 GCGATATTATCCGTTTCCCCGGTTTTGTGAACAATGTTTCCGAAGTGATAAAGGCGCATC 928

Qy         881 CTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTA 940
               | | || || || ||  | |  ||| | |||||||| ||  |    | ||| || || |
Db         929 ATATATTCGTATTGCCGTCTTATTATAGGGAAGGCGTTCCCCGAAGCACCCAGGAGGCAA 988

Qy         941 TGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATG 1000
               ||||  | || ||  | || || || ||  |||| || || || ||||| |    |     |
Db         989 TGGCCGTCGGCAGGGCGGTGATTACGACGGATGTCCCCGGATGCAGGGAAACGGTTGCCG 1048

Qy        1001 ATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGA 1060
               |   |||||| |||||  ||||    || |   |         |  |||| ||||||||||
Db        1049 ACAAGGTCAACGGCTTCCTGATCGAACCTTGGAATCCCCGCATCTTGGCCGAAAAAATGA 1108

Qy        1061 AATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTG 1120
               |||||||| || || |  | |    || | |  |||| |  |||  |       | |||
Db        1109 TTTATTTTATCGAAAACAGGGCTGCCGTCCGCCTGATGGCGAATGCAAGTTATGCGATTG 1168

Qy        1121 CAGAAAAAAACTTTGATGC 1139
               |  || | || || |||||
Db        1169 CCAAAGATAAATTCGATGC 1187


RESULT 7
ABZ41249
ID   ABZ41249 standard; DNA; 1227 BP.
XX
AC   ABZ41249;
XX
DT   07-MAR-2003  (first entry)
XX
DE   N. gonorrhoeae nucleotide sequence SEQ ID 7087.
XX
KW   Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO200279243-A2.
XX
PD   10-OCT-2002.
XX
PF   12-FEB-2002; 2002WO-IB002069.
XX
PR   12-FEB-2001; 2001GB-00003424.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Fontana MR,  Pizza M,  Masignani V,  Monaci E;
XX
DR   WPI; 2003-058415/05.
DR   P-PSDB; ABP80279.
XX
PT   New protein from Neisseria gonorrheae, useful for the manufacture of a
PT   medicament for treating or preventing N. gonorrheae infection.
XX
PS   Disclosure; Page 702; 815pp; English.
XX
CC   The present invention relates to proteins from Neisseria gonorrheae.
CC   Also disclosed are the nucleic acid molecules encoding the proteins and
CC   antibodies that specifically bind to the proteins. The composition
CC   comprising the protein, nucleic acid or antibody is useful for the
CC   manufacture of a medicament for treating or preventing N. gonorrheae
CC   infection, this may be in the form of a vaccine or gene therapy.
CC   Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
CC   molecules of the invention
XX
SQ   Sequence 1227 BP; 328 A; 288 C; 296 G; 315 T; 0 U; 0 Other;

  Query Match             18.2%;  Score 217.8;  DB 10;  Length 1227;
  Best Local Similarity   52.1%;  Pred. No. 2.7e-34;
  Matches  573;  Conservative    0;  Mismatches  502;  Indels   24;  Gaps    3;

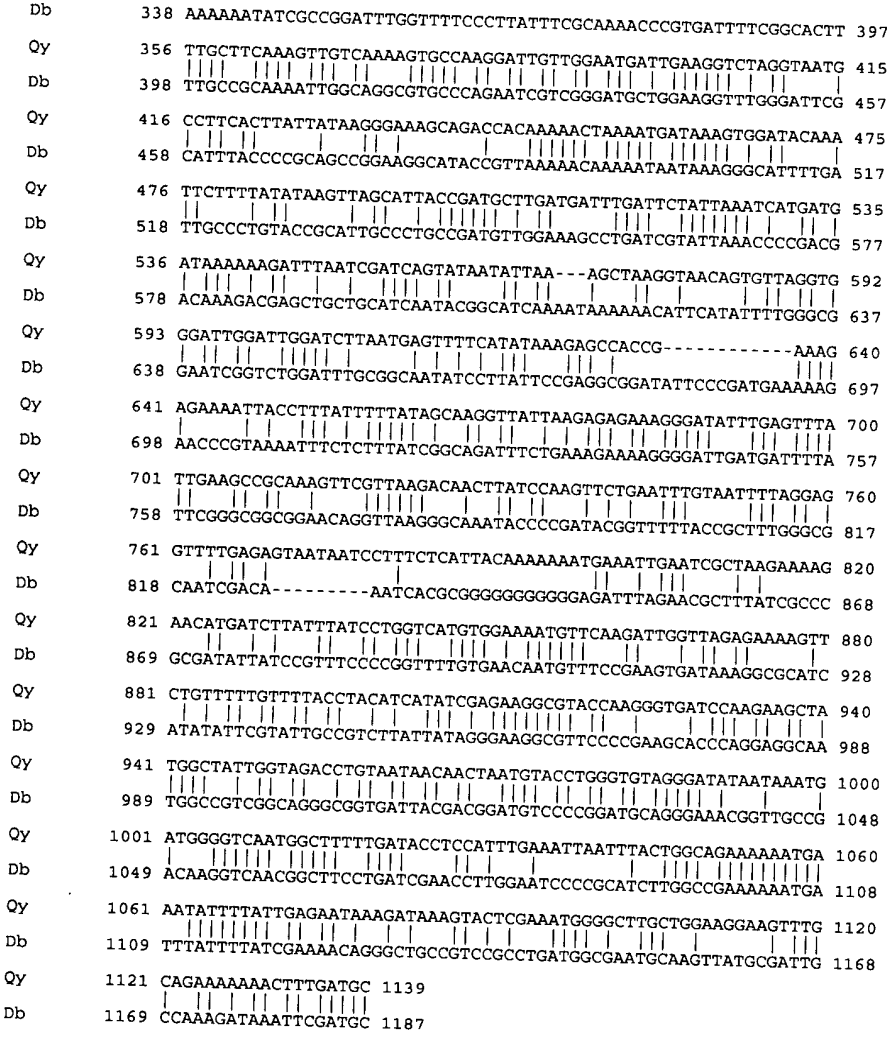
Qy          56 TGAAAAAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAAT 115
               | |  ||||| ||  | ||    |  || || || |  || |   | ||| |
Db          98 TTATGAAAATCGTTTTTATCACAACAGTCGCATCCAGCATTTACGGTTTCCGCGCCCCCG 157

Qy         116 TAATCATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCA 175
               | || |  || ||| | |  ||    || | || |||    | | |   || ||||||
Db         158 TCATTAAAAAATTAATCGGCAAAAACCATCAGGTGTATGCCTTTGTATCGGAGTTTTCCG 217

Qy         176 CTGAAGATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACT 235
                 | | ||  |  | |  |   | |  ||| ||||     ||||  | |      |||
Db         218 ATAATGAGTTGGACATTATCAGGGAAATGGGGGTTACACCCGTTACCTACCGGTCAAACC 277

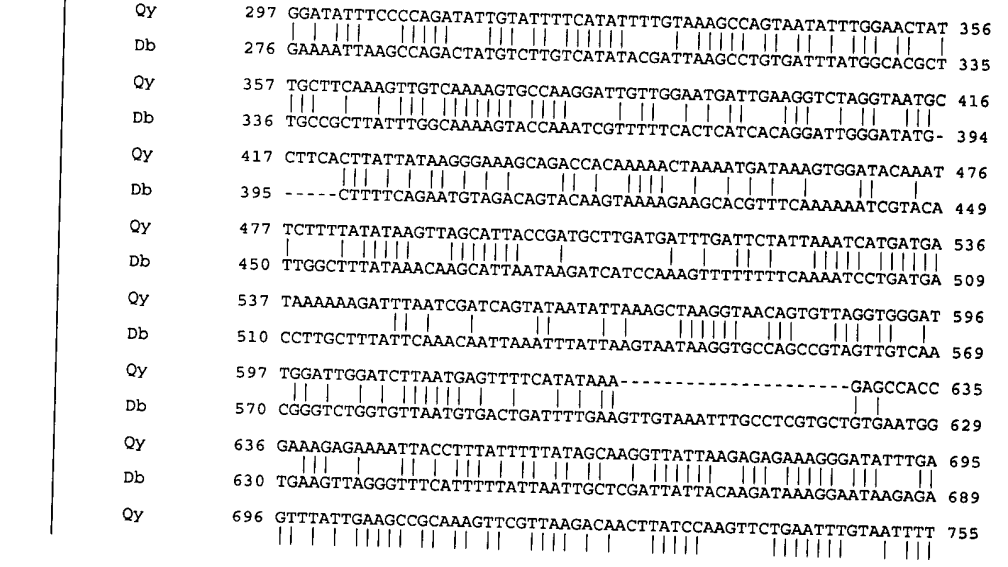
Qy         236 CAAAGGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTA 295
                 |  ||  | || || |||   |||||||   |     |       ||   |||   || |
Db         278 GCAGCGGGGTAAACCCGTTTTCCGATATAAAATCCACCTTCCTCATATTTAAAGCACTCA 337

Qy         296 AGGATATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTA 355
               |  | || || || ||| | || ||  | |||||  | ||| || || || || || |||
```



RESULT 8
ADA29503
ID ADA29503 standard; DNA; 1173 BP.
XX
AC ADA29503;
XX
DT 20-NOV-2003 (first entry)
XX
DE DNA encoding Acinetobacter baumannii protein #790.
XX
KW ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
KW vaccine; plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
DR WPI; 2003-576092/54.
DR P-PSDB; ADA33629.
XX
PT New Acinetobacter baumanii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
PS Example; SEQ ID NO 790; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents DNA encoding an A. baumannii
CC protein.
XX
SQ Sequence 1173 BP; 407 A; 167 C; 224 G; 375 T; 0 U; 0 Other;

Query Match 14.2%; Score 170; DB 9; Length 1173;
Best Local Similarity 52.3%; Pred. No. 1.1e-24;
Matches 492; Conservative 0; Mismatches 415; Indels 33; Gaps 4;

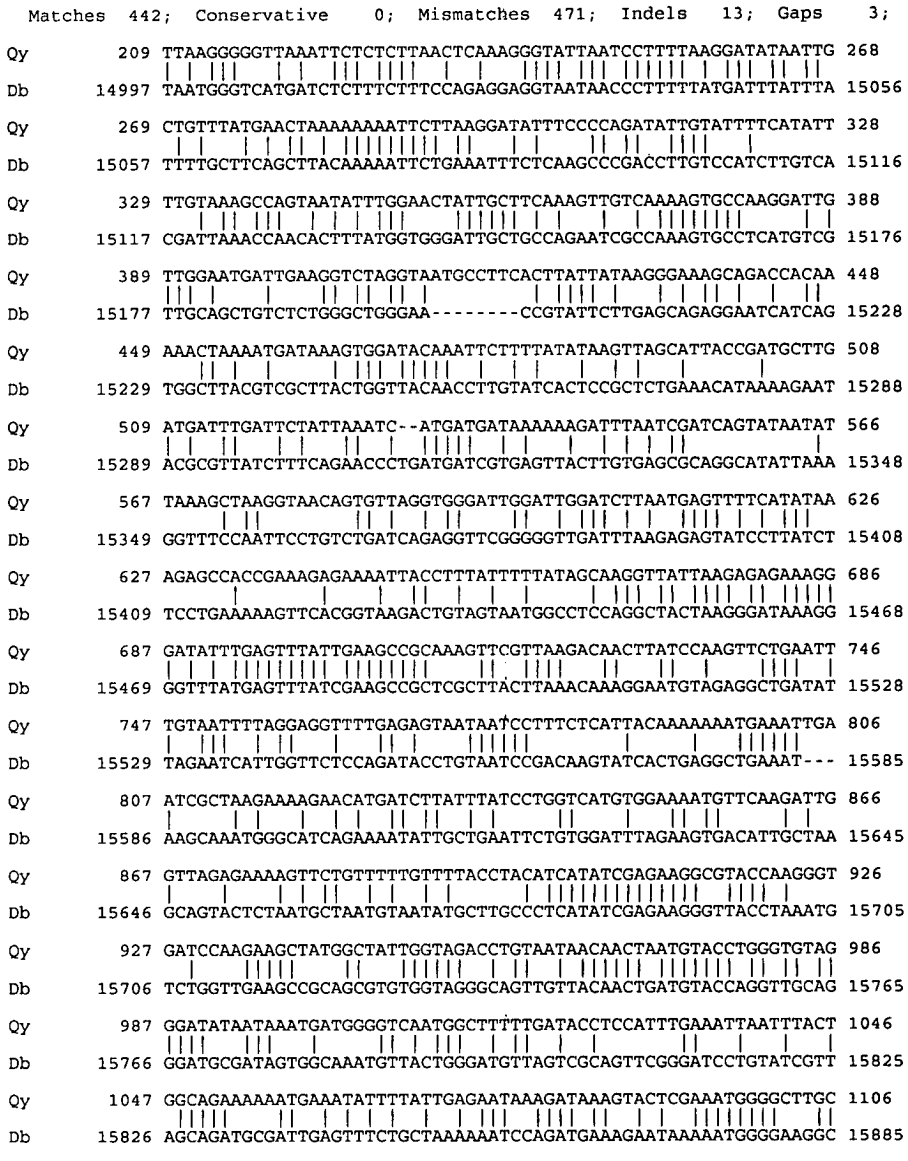




RESULT 9
ADW38769
ID ADW38769 standard; DNA; 17433 BP.
XX
AC ADW38769;
XX
DT 24-MAR-2005 (first entry)
XX
DE Shigella dysenteriae O-antigen DNA SEQ ID NO:1.
XX
KW ds; antigen; O-antigen; gram-neg. bacterium.
XX
OS Shigella dysenteriae.
XX
PN CN1451663-A.
XX
PD 29-OCT-2003.
XX
PF 20-JAN-2003; 2003CN-00100358.
XX
PR 20-JAN-2003; 2003CN-00100358.
XX
PA (UYNA-) UNIV NANKAI.
XX
PI Wang L, Tao J, Feng L;
XX
DR WPI; 2004-110087/12.
XX
PT Nucleotide specific to O-antigen of shigella dysenteriae I and Bacillus
PT coli 0121.
XX
PS Claim 1; SEQ ID NO 1; 39pp; Chinese.
XX
CC The invention relates to a novel nucleotide specific to the O-antigen of
CC Shigella dysenteriae, represented in (ADW38769). The process for
CC obtaining the bacterial O-antigen gene cluster is also included. The O-
CC antigen can be used to detect Shigella dysenteriae.
XX
SQ Sequence 17433 BP; 5366 A; 3017 C; 3662 G; 5388 T; 0 U; 0 Other;

  Query Match   9.7%; Score 116.4; DB 13; Length 17433;
  Best Local Similarity 47.7%; Pred. No. 9.5e-14;

  Matches 442; Conservative   0; Mismatches 471; Indels   13; Gaps   3;



RESULT 10
AEA81303
ID AEA81303 standard; DNA; 14139 BP.
XX
AC AEA81303;
XX

```
DT   11-AUG-2005  (first entry)
XX
DE   E. coli O138 O-antigen gene cluster.
XX
KW   ds; O-antigen; microorganism detection; microorganism identification .
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1137. .2222
FT                   /*tag=  a
FT                   /label= orf1
FT   CDS             2222. .3121
FT                   /*tag=  b
FT                   /label= orf2
FT   CDS             3179. .4057
FT                   /*tag=  c
FT                   /label= orf3
FT   CDS             4062. .4598
FT                   /*tag=  d
FT                   /label= orf4
FT   CDS             4632. .5909
FT                   /*tag=  e
FT                   /label= orf5
FT   CDS             5899. .6855
FT                   /*tag=  f
FT                   /label= orf6
FT   CDS             6848. .7924
FT                   /*tag=  g
FT                   /label= orf7
FT   CDS             7985. .9262
FT                   /*tag=  h
FT                   /label= orf8
FT   CDS             9264. .10289
FT                   /*tag=  i
FT                   /label= orf9
FT   CDS             10333. .11484
FT                   /*tag=  j
FT                   /label= orf10
FT   CDS             11469. .12680
FT                   /*tag=  k
FT                   /label= orf11
XX
PN   CN1554753-A.
XX
PD   15-DEC-2004.
XX
PF   22-DEC-2003; 2003CN-01117855.
XX
PR   22-DEC-2003; 2003CN-01117855.
XX
PA   (UYNA-) UNIV NANKAI.
XX
PI   Wang L,  Kong Q,  Guo H;
XX
DR   WPI; 2005-215821/23.
XX
PT   New O-antigen specific nucleotide of Escherichia coli 0138 type for
PT   detecting and identifying the Escherichia coli O138 in a human body and
PT   in an environment.
XX
PS   Claim 1; SEQ ID NO 1; 31pp; Chinese.
XX
CC   This invention describes the novel O-antigen gene cluster isolated from
CC   Escherichia coli O138. The O-antigen gene cluster has been isolated using
CC   the PCR primers AEA81304 and AEA81305. These primers could also be used
CC   to detect and identify the Escherichia coli O138 O-antigen gene cluster
CC   in humans or in the environment. This sequence represents the E. coli
CC   0138 gene cluster described in the invention.
XX
SQ   Sequence 14139 BP; 4224 A; 2400 C; 3076 G; 4439 T; 0 U; 0 Other;

  Query Match                 9.0%;  Score 107.2;  DB 14;  Length 14139;
```

```
  Best Local Similarity   48.2%;  Pred. No. 6.5e-12;
  Matches  433;  Conservative    0;  Mismatches  448;  Indels   17;  Gaps    4;

Qy        241 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 300
Db      10513 GGGATTTCAATATTAAACGAAATTAATACATGCTTTCATATTTACAACACTGTAAAGAAA 10572

Qy        301 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 360
Db      10573 ATCAAACCCGATATTGTTCATATGGTGACTGTAAAGCCAGTAGCTTATGGTGGTTTGGTA 10632

Qy        361 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 420
Db      10633 TGTAAGCTACTGAAGATAAAGAAACGAGTTGCTTCTATTTCTGGCTTAGGTTATGCATTC 10692

Qy        421 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 480
Db      10693 ATTGACCAGAGCATTAAAGCGAAAATAATTAAAAATATAGTCGTTTTACTCTACAGAATT 10752

Qy        481 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 540
Db      10753 TCTCTTAAAAATAAAAATACAAGCGTTATTTTTCAGAATTTAAATGATAAAGATTATTTT 10812

Qy        541 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 600
Db      10813 ATTAATTGTG-------GTATAATAAAAAAGCATCAATCCGTTTTAATCAGAGGTTCTGG 10865

Qy        601 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 660
Db      10866 TGTTGATCTAGTGAGATATAAGGTGAAGCCGGAGCCATTAGGTAAACCAGTTGTAATGTT 10925

Qy        661 A-TAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGT 719
Db      10926 AGTATCTCGACTATTATATGATAAAGGTATCCGTGAGTTTATAGATGCAGTGCGTTTAGT 10985

Qy        720 TAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCC 779
Db      10986 TAAACAAAAAATTGATTTTTCCG---CTGTGCTTGTAGGAAGTATTGATGCAAATCCAAA 11042

Qy        780 TTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCC 839
Db      11043 CTCTGCAAAAGAAAATGAAATAATGCAATGGCAAAATGAAGGAATTATTGAGTATTGGGG 11102

Qy        840 TGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTAC 899
Db      11103 ATATCGTGATGACATTCCAGAAGTCTTGTGCCAG------TCTAATCTTGTCGTTTTGCC 11156

Qy        900 ATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGT 959
Db      11157 CTCATATCGTGAAGGTCTTCCAAAATGTTTGATTGAAGCGGCAGCTTGTGGTCGGGCAGT 11216

Qy        960 AATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTT 1019
Db      11217 AGTGACTACGGATGTGCCTGGGTGTCGTGATGCTATTATACCAAACGTTACAGGAATATT 11276

Qy       1020 GATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAA 1079
Db      11277 GGTTGAAGTTAAAGATCACATAACGTTAGCAAATGCAATAGAGACACTTATTATTAATCC 11336

Qy       1080 AGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGAT 1137
Db      11337 CGAGCTACGGCATCAGATGGCCGTTAAAGGCAGAGAACTTGCAGAACAGGTATTTGAT 11394
```

```
RESULT 11
AEA29363
ID   AEA29363 standard; DNA; 13234 BP.
XX
AC   AEA29363;
XX
DT   28-JUL-2005  (first entry)
XX
DE   Shigella boydii O-antigen gene cluster SEQ ID NO:1.
```

XX
KW O-antigen; gene cluster; ds.
XX
OS Shigella boydii.
XX

FH	Key	Location/Qualifiers
FT	CDS	1329. .2222
FT		/*tag= a
FT		/product= "ORF1"
FT	CDS	2222. .3121
FT		/*tag= b
FT		/product= "ORF2"
FT	CDS	3179. .4063
FT		/*tag= c
FT		/product= "ORF3"
FT	CDS	4256. .5353
FT		/*tag= d
FT		/product= "ORF4"
FT	CDS	5366. .6130
FT		/*tag= e
FT		/product= "ORF5"
FT	CDS	6127. .7311
FT		/*tag= f
FT		/product= "ORF6"
FT	CDS	7305. .7945
FT		/*tag= g
FT		/product= "ORF7"
FT	CDS	8090. .8751
FT		/*tag= h
FT		/product= "ORF8"
FT	CDS	8748. .9620
FT		/*tag= i
FT		/product= "ORF9"
FT	CDS	9583. .10752
FT		/*tag= j
FT		/product= "ORF10"
FT	CDS	10767. .11297
FT		/*tag= k
FT		/product= "ORF11"
FT	CDS	11330. .12436
FT		/*tag= l
FT		/product= "ORF12"

XX
PN CN1546508-A.
XX
PD 17-NOV-2004.
XX
PF 03-DEC-2003; 2003CN-01107158.
XX
PR 03-DEC-2003; 2003CN-01107158.
XX
PA (UYNA-) UNIV NANKAI.
XX
PI Feng L, Yang J, Peng X;
XX
DR WPI; 2005-215504/23.
XX
PT Nucleotide specific for Shigella boydii 2 O-antigen useful for detecting
PT and identifying Shigella boydii 2 in human body and environment.
XX
PS Claim 1; SEQ ID NO 1; 33pp; Chinese.
XX
CC The invention relates to a nucleotide specific for Shigella boydii 2 O-
CC antigen, which is the total nucleotide sequence of the gene cluster for
CC controlling the synthesis of O-antigens in Shigella boydii 2, e.g.
CC isolated nucleotide represented by AEA29363, with overall length of 13234
CC bases, or nucleotide of AEA29363 including one or more inserted, deleted
CC or substituted bases and sustaining the functions of the isolated
CC nucleotides. Also described are oligonucleotides of the wzx genes in the
CC O-antigen gene cluster originating from the Shigella boydii 2, and genes
CC and wzy genes having similar functions with wzx, and genes having similar
CC functions with wzy genes or wzy genes. The invention proves the high
CC specificity of oligonucleotides for Shigella boydii 2 O-antigen through
CC PCR. A method for detecting and identifying Shigella boydii 2 in human
CC body and the environment by means of the oligonucleotide according to the
CC invention is also disclosed.
XX
SQ Sequence 13234 BP; 4191 A; 2065 C; 2675 G; 4303 T; 0 U; 0 Other;

Query Match 8.8%; Score 105; DB 14; Length 13234;
Best Local Similarity 47.4%; Pred. No. 1.8e-11;
Matches 457; Conservative 0; Mismatches 490; Indels 18; Gaps 4;

```
Qy      177 TGAAGATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTC 236
            |||| ||    |    |||   ||  |||| |||    |  ||  |  ||||||  |
Db     9741 TGAATATAAGGATTCTATTTTAAAATTCGGCGCTAAATTTGAGAAGGTTGATCTTAAACC 9800

Qy      237 AAAGGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAA 296
            |||   |||||||  |||||  ||   |  ||  |     |    |  |||||| | ||
Db     9801 AAAATCTATTAATATTTTTACTGAATGTACATCTCTAATATCATATTATAAAATTATAAA 9860

Qy      297 GGATATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTAT 356
            |  |   ||| |  || |||| | | |     ||  |   |||| | |   | |
Db     9861 AAAATTCAATCCACAGGTTATATTAACGTTTACACCCAAAGCTAATATATATTGTGGTTT 9920

Qy      357 TGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGC 416
            | | |   |  | | |  |||  |  |  |  |||   || || || | |||
Db     9921 GCTTGCTTATAAATTTGACGCGAAAGTGATACCT---AATATTTCTGGCCTTGGAAGTGC 9977

Qy      417 CTTCACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAAT 476
            ||  | ||  ||      | || | |  || | ||||  | |||    || | |
Db     9978 ATTTGTTAATAGTA---------ATACGATATTAAGTCAAATTGTTAAGCTATTATATAA 10028

Qy      477 TCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGA 536
            |  ||  ||| |||| |   |  || | |                |||
Db     10029 GATAGCATTAAAGAAAGCAGCATTCGTTTTTTTTCAAAACGAGGATGATCGTCAGTTATT 10088

Qy      537 TAAAAAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGAT 596
            |  ||  |  ||  | ||| | | |  ||  | ||  ||||
Db     10089 GATTAATTCTGGCTGCGTCAATTATGAGAAAACATGCCGAATATTTGGGTCTGGTGTTGA 10148

Qy      597 TGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTAT 656
            |  ||  |  ||  |  |  | ||  |  | |  |  | | ||||
Db     10149 TTTATCAAAATTTCTTCCCTCAAATAAGAATTTAAAGCAAAGATCATGTGTGAAATTCAT 10208

Qy      657 TTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTT 716
            |  | ||  ||  |  ||      |||| ||| || | | | || | |||||| |  |
Db     10209 ACTTGTTGCTAGACTTCTATATACCAAAGGCATACTTCACTATCTTAAGGCCGCAGAAAT 10268

Qy      717 CGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAA 776
            | ||  |  |||||  | ||  |  |||| | | |||||  | |||||  | ||| |
Db     10269 AATAAAATCCAAATATCCTAATTGTTCATTTTCACTGTTAGGTGCATTTGAAAAAAATGA 10328

Qy      777 TCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTA 836
            |   | |   ||  |   ||||||  |  | |  |  |  |||  |||
Db     10329 AAAGAT---AATCACGAAGGAATTGATTGAAGGTTTTTGTACGAGGGGTATAGTTAATTA 10385

Qy      837 TCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACC 896
            |||  |  ||  | |||  |  |  |  ||||  |  ||  | || |||||
Db     10386 CTATGGGACTTCCGATGACGTTGCTAGTATAATGAAGCATTATGATGCAATAGTATTACC 10445

Qy      897 TACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACC 956
            |||  ||| ||||||| |||||  |  | |  |  |||| |   ||| |||  | |
Db     10446 ATCATTTTATAGAGAAGGTGTACCTAAAGCGTTGATTGAAGGTGCTTCTAGTGGGTTAGC 10505

Qy      957 TGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTT 1016
            | |  | |||||  ||  |||  |||||||||||  |  |  |||  || ||||| ||
Db     10506 TATTTTGACAACAGATAATGTTGGCTGTAGGGATATGGTGCAAAATGAAGTTAATGGTTT 10565

Qy     1017 TTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAA 1076
            ||||  |  ||  |||  || ||   |  || |  ||  | | | ||||
Db     10566 TTTGTGTCAACCTAATGATCTTGATAGCTTAGTTGACATTATTGAAAAGTATATTCTACT 10625

Qy     1077 TAAAGATAAAGT---ACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTT 1133
            | |||  ||  |   |  |  ||||    ||   ||  | ||||| |||  | | |
Db     10626 GACAGACGAAATGAAAAACACAATGTCTATTAAGTCGAGAGAATTTGCTGAATCACATTG 10685
```

```
Qy       1134 TGATG 1138
              |||||
Db      10686 TGATG 10690


RESULT 12
AEC31411
ID   AEC31411 standard; DNA; 11717 BP.
XX
AC   AEC31411;
XX
DT   20-OCT-2005  (first entry)
XX
DE   E. coli type O100 O-antigen gene cluster.
XX
KW   O-antigen; gene expression; DNA detection; gene; ds.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1108. .2190
FT                   /*tag=  a
FT                   /note= "Orf1"
FT   CDS             2193. .3092
FT                   /*tag=  b
FT                   /note= "Orf2"
FT   CDS             3145. .4023
FT                   /*tag=  c
FT                   /note= "Orf3"
FT   CDS             4027. .4575
FT                   /*tag=  d
FT                   /note= "Orf4"
FT   CDS             4577. .4981
FT                   /*tag=  e
FT                   /note= "Orf5"
FT   CDS             4994. .6235
FT                   /*tag=  f
FT                   /note= "Orf6"
FT   CDS             6219. .7415
FT                   /*tag=  g
FT                   /note= "Orf7"
FT   CDS             7449. .8618
FT                   /*tag=  h
FT                   /note= "Orf8"
FT   CDS             8627. .9568
FT                   /*tag=  i
FT                   /note= "Orf9. Specifically claimed in Claim 3"
FT   CDS             9584. .10723
FT                   /*tag=  j
FT                   /note= "Orf10. Specifically claimed in Claim 3"
XX
PN   CN1563051-A.
XX
PD   12-JAN-2005.
XX
PF   19-APR-2004; 2004CN-00019036.
XX
PR   19-APR-2004; 2004CN-00019036.
XX
PA   (TIAN-) TIANJIN BIOCHIP TECH CO LTD.
XX
PI   Wang L,  Li Y,  Feng L;
XX
DR   WPI; 2005-307126/32.
XX
PT   Nucleotide peculiar to 0-antigen of 0100 type bacillus coli.
XX
PS   Claim 1; SEQ ID NO 1; 35pp; Chinese.
XX
CC   The invention relates to a nucleotide sequence (SEQ ID No:1) which is
CC   specific to the O-antigen of Escherichia coli O100. The nucleotide
CC   sequence represents the complete sequence (11717bp) of a gene cluster for
CC   controlling the synthesis of O-antigen in Escherichia coli O100. The
CC   nucleotide sequence can have one or several inserted, deleted or
CC   substituted bases, whilst still retaining its function. Also described is
CC   a PCR primer for the glycosyltransferase (galF) gene and a PCR primer for
CC   the oligosaccharide unit treatment (gnd) gene, both of which are
CC   contained within the E. coli type O100 O-antigen gene cluster. The
CC   invention also discloses a method for detecting and identifying E. coli
CC   type O100 in humans. This sequence represents the E. coli type O100 O-
CC   antigen gene cluster of the invention.
XX
SQ   Sequence 11717 BP; 3382 A; 2158 C; 2580 G; 3597 T; 0 U; 0 Other;

  Query Match              8.7%;  Score 103.4;  DB 14;  Length 11717;
  Best Local Similarity   46.8%;  Pred. No. 3.7e-11;
  Matches  440;  Conservative    0;  Mismatches  486;  Indels   15;  Gaps    3;

Qy       201 ATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTTAAGGA 260
            || ||| | | | ||| | | || || | |||| |||||||||| ||
Db      9739 ATTAGGCTGTGATGTTGTTGATATTCCTATTTCATCGCAAGGTAAAAATCCTTTTATTGA 9798

Qy       261 TATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGATATTTCCCCAGATATTGTATT 320
            || | | || ||| | | || ||| ||| | || | | | | | ||
Db      9799 CATGCTACTTATAAATCGACTGAGAGTTGTTTATAAAAGGATTAAACCGGATTTTGTTTT 9858

Qy       321 TTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCTTCAAAGTTGTCAAAAGTGCC 380
            | ||| | |||||||| |||| ||| | ||||||| | ||||| | |
Db      9859 TCACTATACTATAAAGCCAAATATATATGGCTCGATTGCTGCTGGCTTGTCTCGA---AC 9915

Qy       381 AAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTCACTTATTATAAGGGAAAGCA 440
            || | | || | | || || ||| ||| |||| | | |
Db      9916 AAAATCAATAGCTATTACTACCGGGCTTGGTTATGTTTTCAATAACTCTTCATTTTTCAC 9975

Qy       441 GACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACC 500
            || || | | || | | | | | | | | | | | | | | | | ||
Db      9976 TCACATAATTTCACATTTGTACTGCTTTGCATTTAAATTCGCAAAAGAAGTCTGGTTTTT 10035

Qy       501 GATGCTTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTA 560
            || |||||| | | | || | |||| || | | | | | | |
Db     10036 GAATGAAGATGATAGAAATGTGTTTATAC---------AAAGAGGTATTATCTCACCTGG 10086

Qy       561 TAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTC 620
            || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     10087 CAAGACTCATATTCTTGATGGCGAAGGTGTCGATACTAGTTTCTTCATTCCTAGGGCAAA 10146

Qy       621 ATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGA 680
            | || || ||| ||| |||| ||| | | | |||||| | | | | ||
Db     10147 ACCTATTTCGCATAAAAAATCAAAAGTTACGTTTTTGCTCGTTGCAAGGATGCTGTGGGA 10206

Qy       681 GAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTC 740
            ||||| | | |||| || | | | ||| | || | | | | | | |
Db     10207 CAAAGGTGTTGGGATATACGTTGATGCGGCTCGTAAACTAAAAGAACAGTACCCAGATAC 10266

Qy       741 TGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGA 800
            ||||||| | ||||| | | || ||||||| | | | | | | ||
Db     10267 TGAATTTCAGTTGTTAGGGGCCTGCGATGTCGATAATCCGAGCGCAATTCCACGTGATAT 10326

Qy       801 AATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCA 860
            || ||  |||| || ||| |||| ||  | | |||| | |
Db     10327 TATAGATAATTGGCATGAAGAGGGTGTAGTTAATTATTTGGGAATGACGAATGATGTGCG 10386

Qy       861 AGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACC 920
            ||  ||  | ||| || ||| | | ||||| || || || | | | |
Db     10387 CACCGTTGTATCCCAAGCCGACTGTAATGTATTGCCTTCTTATTATCGCGAGGGAGTTCC 10446

Qy       921 AAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGG 980
            | |  |||||  | || || | ||| ||| | || ||  ||  ||
Db     10447 TCGAACATTGATGGAAGCAGCATCAATGGGCAAACCAGTACTTACGACCGATAATGTCGG 10506

Qy       981 GTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAA 1040
            ||||| || | | | || | | | || ||||| || || | | | |
Db     10507 ATGTAGAAATGTTGTTATAGAAGCGCAAACAGGTTTTTTATGCAAGACAAAAGATGTCGA 10566

Qy      1041 TTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAAGTAC---TCGAAAT 1097
```

```
Db      10567 TAGTTTAGTTGACGCTATGAATAAGATTCTCCATTTAACCCCTGAGGAACGTAGCGAAAT 10626

Qy       1098 GGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATG 1138

Db      10627 GGGTCTAAATGGTCGGGATTATATGATAAGCCGCTTCGATG 10667
```

RESULT 13
ADF77651
ID ADF77651 standard; DNA; 17623 BP.
XX
AC ADF77651;
XX
DT 26-FEB-2004 (first entry)
XX
DE Francisella tularensis strain Schu 24 genomic DNA SEQ ID NO:1.
XX
KW enzyme; immunogenic moiety; immune response;
KW Francisella tularensis infection; antibacterial; vaccine; gene; ds.
XX
OS Francisella tularensis.
XX
PN WO2003102191-A1.
XX
PD 11-DEC-2003.
XX
PF 30-MAY-2003; 2003WO-GB002338.
XX
PR 31-MAY-2002; 2002GB-00012666.
XX
PA (MINA) UK SEC FOR DEFENCE.
XX
PI Prior JL, Prior RG, Hitchen PG, Dell A, Titball RW;
XX
DR WPI; 2004-043118/04.
DR P-PSDB; ADF77653, ADF77654, ADF77655, ADF77656, ADF77657, ADF77658,
DR ADF77659, ADF77660, ADF77661, ADF77662, ADF77663, ADF77664, ADF77665,
DR ADF77666, ADF77667.
XX
PT New nucleic acids encoding enzymes that produce the O-antigen of
PT Francisella tularensis, useful for diagnosing Francisella tularensis
PT infection, and producing prophylactic or therapeutic vaccines against the
PT infection.
XX
PS Claim 3; SEQ ID NO 1; 48pp; English.
XX
CC The present invention describes a nucleic acid (I) which encodes a series
CC of enzymes or enzyme fragments, which when expressed together in a cell
CC are able to produce an immunogenic moiety that produces an immune
CC response in animal to which it is administered, where the response is
CC protective against Francisella tularensis infection. Also described: (1)
CC a nucleic acid comprising a fragment of the genomic sequence of F.
CC tularensis of 17623 bp (ADF77651) (II), which may be used to detect the
CC presence of (II) in a sample; (2) an expression vector comprising (I);
CC (3) a host cell transformed with the vector; (4) a method of preparing a
CC prophylactic or therapeutic vaccine comprising transforming a host cell
CC with (I), culturing the host cell and recovering a protective immunogenic
CC moiety from the culture; (5) recombinant O-antigen of F. tularensis
CC obtained by the method of (4); (6) a live vaccine vector comprising (I);
CC (7) a vaccine comprising the live vaccine vector or the recombinant O-
CC antigen, and a pharmaceutical carrier; (8) a method of diagnosing
CC infection by F. tularensis comprising detecting (I) in a sample from a
CC patient suspected of having the infection; (9) a method of
CC differentiating between strains of F. tularensis comprising selecting
CC primers or probes which are specific for (II), and not for similar
CC sequences in subspecies other than F. tularensis subspecies tularensis,
CC or which produce distinguishable products when used to analyse other
CC species, and conducting an analysis using the primers or probes; (10)
CC lipopolysaccharide (LPS) obtained from F. tularensis subspecies
CC tularensis for use as a vaccine against infection by F. tularensis; and
CC (11) a pharmaceutical composition comprising LPS and a pharmaceutical
CC carrier. (I) has antibacterial activity, and can be used in vaccines. The
CC nucleic acids can be used for diagnosing F. tularensis infection, and for
CC the production of the O-antigen of F. tularensis for producing
CC prophylactic or therapeutic vaccines against F. tularensis infection. The
CC present sequence is used in the exemplification of the present invention.
XX
SQ Sequence 17623 BP; 6049 A; 2073 C; 3356 G; 6145 T; 0 U; 0 Other;

Query Match 7.4%; Score 88.4; DB 12; Length 17623;
Best Local Similarity 49.4%; Pred. No. 4e-08;
Matches 310; Conservative 0; Mismatches 291; Indels 27; Gaps 2;

```
Qy       457 ATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTG 516

Db      3624 ATTGTTCAGAAGTTTATAATATCTTTATATAAGTTATCATTTAAAAGCACCACAAAAGTA 3683

Qy       517 ATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAAAGCTAAG 576

Db      3684 TTCTTTCAGAATGAGCAAAATAAAAAGTTATTTATAGCTAAGAAAATAATCAGTGGAGAA 3743

Qy       577 GTAACAGTGTTA---------GGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAA 627

Db      3744 AAATCAATATTATTACCAGGTTCTGGGGTAAACTTAGATGAAAATAAATATGTTGACTAT 3803

Qy       628 GAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGG 687

Db      3804 CCTAAAGACCAAGGAATATTAAAATTCGTTTTTCTTGGCCGAATAATGAAAGAAAAGGGG 3863

Qy       688 ATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTT 747

Db      3864 ATTTATGAATTGTTAGAAGCCTTTGCTATACTTGAGAAAAAATATAAAAATATTAGTCTT 3923

Qy       748 GTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAA 807

Db      3924 GACATTTATGGTTTTTGTGATGAAAATAAATCTAATTTTATGGGAAAGGTTAATACGATA 3983

Qy       808 TCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGG 867

Db      3984 AAATCAGTAAAATTTTATGGTTTTACTGATAATACTAAAGAAAAAATAGCT--------- 4034

Qy       868 TTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTG 927

Db      4035 ---------AGTGCACATGCAGTTGTTTTGCCATCTTACCATGAAGGAATGTCAAATGTG 4085

Qy       928 ATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGG 987

Db      4086 CTGTTAGAAGCAGCTGCGATAGGTAGACCTGTAATTGCGTCAGATATTCCTGGGTGTAGA 4145

Qy       988 GATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTG 1047

Db      4146 GAAATTTTTGATGATGGTCTCTCTGGCTTATCATGTAACCCTAATGATGTGAGTTCTTTA 4205

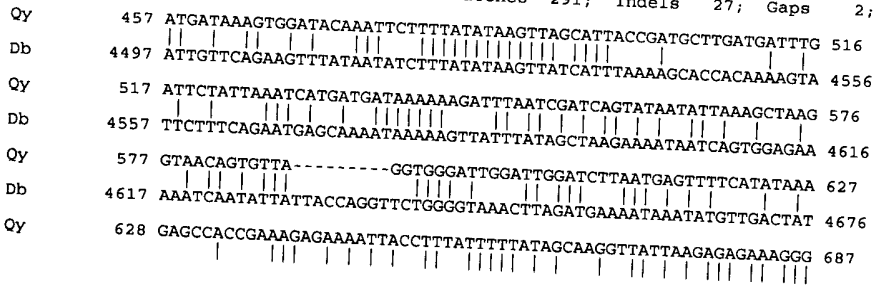
Qy      1048 GCAGAAAAAATGAAATATTTTATTGAGA 1075

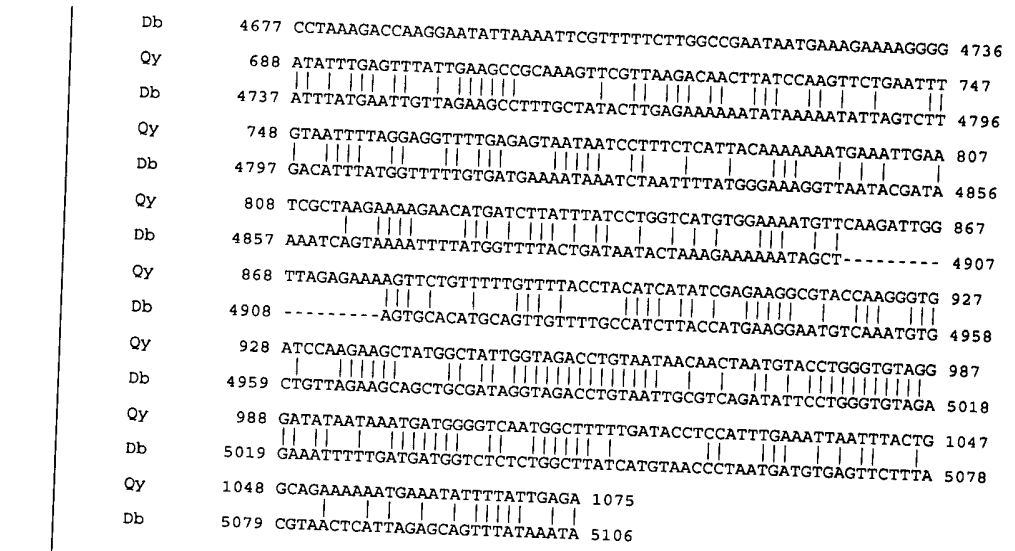
Db      4206 CGTAACTCATTAGAGCAGTTTATAAATA 4233
```

RESULT 14
ADF77691
ID ADF77691 standard; DNA; 22401 BP.
XX
AC ADF77691;
XX
DT 26-FEB-2004 (first entry)
XX
DE Francisella tularensis strain Schu 24 genomic DNA SEQ ID NO:41.
XX
KW enzyme; immunogenic moiety; immune response;
KW Francisella tularensis infection; antibacterial; vaccine; gene; ds.
XX
OS Francisella tularensis.
XX
PN WO2003102191-A1.
XX

PD 11-DEC-2003.
XX
PF 30-MAY-2003; 2003WO-GB002338.
XX
PR 31-MAY-2002; 2002GB-00012666.
XX
PA (MINA) UK SEC FOR DEFENCE.
XX
PI Prior JL, Prior RG, Hitchen PG, Dell A, Titball RW;
XX
DR WPI; 2004-043118/04.
DR P-PSDB; ADF77652, ADF77653, ADF77654, ADF77655, ADF77656, ADF77657,
DR ADF77658, ADF77659, ADF77660, ADF77661, ADF77662, ADF77663, ADF77664,
DR ADF77665, ADF77666, ADF77667, ADF77668, ADF77669, ADF77670.
XX
PT New nucleic acids encoding enzymes that produce the O-antigen of
PT Francisella tularensis, useful for diagnosing Francisella tularensis
PT infection, and producing prophylactic or therapeutic vaccines against the
PT infection.
XX
PS Disclosure; SEQ ID NO 41; 48pp; English.
XX
CC The present invention describes a nucleic acid (I) which encodes a series
CC of enzymes or enzyme fragments, which when expressed together in a cell
CC are able to produce an immunogenic moiety that produces an immune
CC response in animal to which it is administered, where the response is
CC protective against Francisella tularensis infection. Also described: (1)
CC a nucleic acid comprising a fragment of the genomic sequence of F.
CC tularensis of 17623 bp (ADF77651) (II), which may be used to detect the
CC presence of (II) in a sample; (2) an expression vector comprising (I);
CC (3) a host cell transformed with the vector; (4) a method of preparing a
CC prophylactic or therapeutic vaccine comprising transforming a host cell
CC with (I), culturing the host cell and recovering a protective immunogenic
CC moiety from the culture; (5) recombinant O-antigen of F. tularensis
CC obtained by the method of (4); (6) a live vaccine vector comprising (I);
CC (7) a vaccine comprising the live vaccine vector or the recombinant O-
CC antigen, and a pharmaceutical carrier; (8) a method of diagnosing
CC infection by F. tularensis comprising detecting (I) in a sample from a
CC patient suspected of having the infection; (9) a method of
CC differentiating between strains of F. tularensis comprising selecting
CC primers or probes which are specific for (II), and not for similar
CC sequences in subspecies other than F. tularensis subspecies tularensis,
CC or which produce distinguishable products when used to analyse other
CC species, and conducting an analysis using the primers or probes; (10)
CC lipopolysaccharide (LPS) obtained from F. tularensis subspecies
CC tularensis for use as a vaccine against infection by F. tularensis; and
CC (11) a pharmaceutical composition comprising LPS and a pharmaceutical
CC carrier. (I) has antibacterial activity, and can be used in vaccines. The
CC nucleic acids can be used for diagnosing F. tularensis infection, and for
CC the production of the O-antigen of F. tularensis for producing
CC prophylactic or therapeutic vaccines against F. tularensis infection. The
CC present sequence is used in the exemplification of the present invention.
XX
SQ Sequence 22401 BP; 7754 A; 2745 C; 4259 G; 7643 T; 0 U; 0 Other;

Query Match 7.4%; Score 88.4; DB 12; Length 22401;
Best Local Similarity 49.4%; Pred. No. 4.2e-08;
Matches 310; Conservative 0; Mismatches 291; Indels 27; Gaps 2;





RESULT 15
ADW00484
ID ADW00484 standard; DNA; 18900 BP.
XX
AC ADW00484;
XX
DT 24-FEB-2005 (first entry)
XX
DE Escherichia coli O52 O-antigen gene cluster partial sequence, SEQ ID 1.
XX
KW O-antigen; gene; ds.
XX
OS Escherichia coli; O52.
XX
PN CN1462801-A.
XX
PD 24-DEC-2003.
XX
PF 27-MAY-2003; 2003CN-00129936.
XX
PR 27-MAY-2003; 2003CN-00129936.
XX
PA (UYNA-) UNIV NANKAI.
XX
PI Wang L, Yang J, Feng L;
XX
DR WPI; 2004-248670/24.
DR GENBANK; AY528413.
XX
PT Nucleotide specific to O antigen of O52 colibacillus.
XX
PS Claim 1; SEQ ID NO 1; 35pp; Chinese.
XX
CC The present invention relates to the nucleotide sequence (I; ADW00484)
CC specific to the O-antigen of Escherichia coli 052 (II). (I) is a complete
CC nucleotide sequence of the genome controlling the synthesis of (II),
CC including the separated (I), (I) with one or more inserted, deficient, or
CC substituted bases. Also disclosed is an oligonucleotide of
CC glycocosyntransferase gene and oligosaccharide unit treating gene in O-
CC antigen genome. It has high specificity to said O-antigen. A process for
CC detecting Escherichia coli 052 from human body and environment is also
CC disclosed.

XX
SQ Sequence 18900 BP; 5819 A; 3244 C; 4029 G; 5808 T; 0 U; 0 Other;

Query Match 6.6%; Score 78.4; DB 13; Length 18900;
Best Local Similarity 45.2%; Pred. No. 4.2e-06;
Matches 452; Conservative 0; Mismatches 536; Indels 12; Gaps 4;

```
Qy        138 AGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAAGATCTTAAAGTACTTTC 197
              |||  || |||| |||  | ' || ||| ||| |  ||       |||  |     || |
Db      16330 AGGATATGATGTCTATATTTTTGCGAATTATTCTGACAAGTCTATTCTGGATCGCTTTAC 16389

Qy        198 GTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTTAA 257
                |  ||  ||||| |  || |         | ||     |   ||| ||||||  ||
Db      16390 TAGTTTAGGATTTAAGTGTATTGATAGTAAAATAAAAGAGCAAAATATCAATCCTGTTGT 16449

Qy        258 GGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGATATTTCCCCAGATATTGT 317
                |    | | |   |     |   | ||   | ||  | |||   || ||||||||
Db      16450 TTTCTTTTGCGATATAACTCGCTCATTTAGAACATTGAATAAAATTAATCCTGATATTGT 16509

Qy        318 ATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCTTCAAAGTTGTCAAAAGT 377
              |          | || || |   |  || |   | |     | |   |   |||
Db      16510 GCATAGCCGCACGATTAAACCTGGTGTGATTAGTTGTTTATGGGCCAGAATCCGAAA--- 16566

Qy        378 GCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTCACTTATTATAAGGGAAA 437
                ||  || ||    |   |||  ||  ||||||  |  ||        ||||||  |
Db      16567 CAAAAAAATGGTCTATAGTTTTGTTGGGTTAGGTAGAGTTTTTGAAAGCAATAAGGTCA- 16625

Qy        438 GCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATT 497
                || |||     ||      | | | ||  |  | |   | ||   |    | |
Db      16626 ----TTTACCAAATGGTTAAGTTCTTAATTGCTAATATGTATAGAAGATTTTTCTTAAAT 16681

Qy        498 ACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCA 557
              |  |||  |||     |   |     ||||  |       |||||| | || | | || |
Db      16682 ATTGATTGTTGTATCCTCTTTGAACATAAAAAAGATCAGCAAAAAATAATTGAACTATTA 16741

Qy        558 GTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTT 617
              | ||| |   |||  |||      |||  | | |||    ||  |  || |  |  | ||
Db      16742 G-ATATTCCAAAAAATAAAACTGAAGTTATTGATGGCGCAGGTATTAATATAGACTATTT 16800

Qy        618 TTCATATAAAGAGCCACCGAAAGA---GAAAATTACCTTTATTTTTATAGCAAGGTTATT 674
              ||  ||||     ||||| ||  |    ||  |||   || |||||  |    |  | |
Db      16801 TTGTTATAGTACACCACCTAATAATACCAAGGTTAAAGTTTTTTTTGCAAGTCGCATGCT 16860

Qy        675 AAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCC 734
              |    |||||| ||     ||| ||||| ||    |   | | ||   |
Db      16861 TTGGAGTAAAGGGCTACGCACGTTAATTGATGCTAGTAGAATTTTAAAATTACAAGGCAT 16920

Qy        735 AAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAA 794
              | |||  |  || ||   ||   | |||     ||| |||    | |  ||   |
Db      16921 CGAATTTGAGATATTAGTTGCGGGCATTCTTGTTGACAATGATCGCGACGCTATAAGTAT 16980

Qy        795 AAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAAAA 854
              || | | || |||  |     ||     |||| | ||||  |  ||  |       |
Db      16981 AACTCAGATAGAAGAGTGGCATAACTCCGGTGATATCATTTGGCTGGGCAAGCGTAGCGA 17040

Qy        855 TGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAAGG 914
              | |  | || |   | || |       | |   | | ||| ||  |     ||    |||||
Db      17041 TATCAAGGAATTAATTGAAAGTGTGGATATAGTCGCTTTGCCATCTGTTTACTCTGAAGG 17100

Qy        915 CGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAATGT 974
              | ||  |  | |  |   ||||||    || || ||| | ||||| ||||    | |
Db      17101 TATTCCTCGAATTCTTTTAGAAGCCGGTGCAATCGGTCGTCCTGTTATAAGTTTTGACAC 17160

Qy        975 ACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTTGA 1034
                ||| ||| ||  | |||||  ||| ||    || || ||  |  | |||  |    |
Db      17161 TGGTGGTTGTGGGAGTTTAATACTTGACGGTTATAACGGTTTCCTTGTGCCTAAAGGAAA 17220

Qy       1035 AATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAAGTACTCGA 1094
              | ||  || |  |  | ||  |       |||||    ||     | |       |
Db      17221 TGTGAACCTATTTTCTCAGAAGCTTGGTATCCTTATTAGCGATCCTCTTGAGCGCACTAA 17280

Qy       1095 AATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTT 1134
               |||||||    ||  || |||   | |||  |||| | |
```

Db 17281 GATGGGGCAGAATGCCAGAAAGCGAGTAGAGGAAAAGTAT 17320

Search completed: May 10, 2006, 10:59:35
Job time : 563.376 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 11:00:02 ; Search time 3824.15 Seconds
(without alignments)
14608.170 Million cell updates/sec

Title: US-09-423-093B-1_COPY_739_1932
Perfect score: 1194
Sequence: 1 atgttactatgctgcataca.....aatcaaataatgatttttga 1194

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	109.2	9.1	1238	9	BZ565030	BZ565030 pacs2-164
c	2	96.8	8.1	800	9	BZ559683	BZ559683 pacs2-164
c	3	79.6	6.7	1101	10	CNS0039G	AL063921 Drosophil
	4	74.8	6.3	1101	10	CNS00EVL	AL069706 Drosophil
	5	74.2	6.2	761	9	BZ562735	BZ562735 pacs2-164
	6	67.8	5.7	945	9	BZ569481	BZ569481 pacs2-164
	7	66.6	5.6	1101	10	CNS0021J	AL061936 Drosophil
	8	66	5.5	1101	10	CNS0106X	AL098595 Drosophil
	9	65.4	5.5	1091	7	CF934550	CF934550 TrEST-B10
c	10	65.4	5.5	1101	10	CNS00EVL	AL069706 Drosophil
	11	64.6	5.4	920	9	BZ569436	BZ569436 pacs2-164
c	12	63.6	5.3	1272	9	CC264939	CC264939 CH261-19L
	13	62.8	5.3	903	9	BZ569276	BZ569276 pacs2-164
c	14	62.8	5.3	1150	9	BZ567774	BZ567774 pacs2-164
	15	62.4	5.2	1170	10	CL021844	CL021844 CH216-8P1
c	16	61.6	5.2	1489	10	AG350139	AG350139 Mus muscu
	17	61	5.1	1101	10	CNS00EA9	AL068837 Drosophil
c	18	61	5.1	1201	10	CNS0167M	AL106396 Drosophil
c	19	60.4	5.1	1210	10	CG749728	CG749728 P044-1-C0
	20	60.2	5.0	765	10	BX177930	BX177930 Danio rer
c	21	58.8	4.9	986	8	DN581515	DN581515 92992352
c	22	58.8	4.9	1242	10	CL068807	CL068807 CH216-115
c	23	58.6	4.9	1101	10	CNS017KE	AL108152 Drosophil
c	24	58.6	4.9	1356	9	BZ575664	BZ575664 msh2_4567
	25	58.2	4.9	928	10	CNS00DKY	AL071865 Drosophil
	26	58	4.9	1101	10	CNS0100X	AL098379 Drosophil
c	27	57.8	4.8	1027	10	CNS02T50	AL212733 Tetraodon
c	28	57.6	4.8	1592	10	CG750135	CG750135 P044-3-D0
	29	57.4	4.8	1101	10	CNS000D1	AL065414 Drosophil
	30	57.2	4.8	1010	10	CNS010GD	AL098935 Drosophil
c	31	57.2	4.8	1101	10	CNS0039R	AL063932 Drosophil
	32	57	4.8	1180	8	DN701995	DN701995 CLJ39-A03
c	33	56.8	4.8	1094	10	CNS012FZ	AL101513 Drosophil
c	34	56.4	4.7	521	6	CA997531	CA997531 rg18f07.y
	35	56.4	4.7	689	3	BM594308	BM594308 170006874
c	36	56.4	4.7	1210	1	AJ927826	AJ927826 AJ927826
	37	56.2	4.7	1065	1	AJ927054	AJ927054 AJ927054
	38	56.2	4.7	1367	8	DN683496	DN683496 CGX30-C12
	39	55.8	4.7	1101	10	CNS003BD	AL064091 Drosophil
c	40	55.6	4.7	1101	10	CNS017L2	AL108176 Drosophil
	41	55.6	4.7	1146	10	CNS021G2	AL176843 Tetraodon
	42	55.6	4.7	1277	9	CC253231	CC253231 CH261-180
	43	55.4	4.6	1096	1	AJ927821	AJ927821 AJ927821
c	44	55.4	4.6	1120	1	AJ927743	AJ927743 AJ927743
	45	55.2	4.6	650	3	BM594273	BM594273 170006874

ALIGNMENTS

RESULT 1
BZ565030
LOCUS BZ565030 1238 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_5045.x1 pacs2-164 Pseudomonas aeruginosa genomic clone pacs2-164_5045, genomic survey sequence.
ACCESSION BZ565030
VERSION BZ565030.1 GI:27191946
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1238)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .1238
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_5045"
/clone_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun library."
ORIGIN

Query Match 9.1%; Score 109.2; DB 9; Length 1238;
Best Local Similarity 54.8%; Pred. No. 6e-12;
Matches 238; Conservative 0; Mismatches 193; Indels 3; Gaps 1;

```
Qy        592 GGGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACC 651
              || || ||  | ||||| || | ||| |   |     ||   |   |  |||||
Db        220 GGTATAGGCGTCGATCTTGATTATTTTGCCGCGAGTTCTCCCGTGTTGCGGCCTATTACT 279
```

```
Qy        652 TTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCA 711
Db        280 TTCGTGTTTGTCGCTCGTCTTCTCAAGGAGAAAGGTATCTATGATTTTGTGGAAGCCGCG 339

Qy        712 AAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGT 771
Db        340 AGAATCGTTAGAAAACGTAGTCCGCTCGCGCGTTTTCTGGTCTTGGGAGGCGTCGACAG- 398

Qy        772 AATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTT 831
Db        399 --TAATCCAGGAAAGGTTTCACGGGAGGAAATGAACCAGTGGGTAGCCGAAGGCTTGTTG 456

Qy        832 ATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTT 891
Db        457 GAGTGGCCTGGCCATGTGCATGACATCAAGGCATGGGTGGCCAATACGAGTGTATTCGTG 516

Qy        892 TTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGT 951
Db        517 CTGCCCTCCTATTATAGAGAAGGTGTGCCGAGAAGTAGTCAAGAGGCCATGTCCCTGGGG 576

Qy        952 AGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAAT 1011
Db        577 AAGCCTGTGATAACGACTGACTGGACGGGATGTCGTGAAACGGTACTGGATGGGTTCAAT 636

Qy       1012 GGCTTTTTGATACC 1025
Db        637 GGAATTTTGGTACC 650
```

RESULT 2
BZ559683/c

```
LOCUS       BZ559683                 800 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_1748.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_1748, genomic survey sequence.
ACCESSION   BZ559683
VERSION     BZ559683.1  GI:27176354
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .800
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_1748"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN
```

```
  Query Match             8.1%;  Score 96.8;  DB 9;  Length 800;
  Best Local Similarity   53.0%;  Pred. No. 2e-09;
  Matches  231;  Conservative    0;  Mismatches  202;  Indels    3;  Gaps    1;

Qy        592 GGGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACC 651
```

```
Db        486 GGTATAGGCGTCGATCTTGATTATTTTGCCGCGAGTTCTCCCGTGTTGCGGCCTATTACT 427

Qy        652 TTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCA 711
Db        426 TTCGTGTTTGTCGCTCGTCTTCTCAAGGAGAAAGGTATCTATGATTTTGTGGAAGCCGCG 367

Qy        712 AAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGT 771
Db        366 AGAATCGTTAGAAAACGTAGTCCGCTCGCGCGTTTTCTGGTCTTGGCAGGCGTCGACAG- 308

Qy        772 AATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTT 831
Db        307 --TAATCCAGGAAAGGTTTCACGGGAGGAAATGAACCAGTGGGTAGCCGAAGGCTTGTTG 250

Qy        832 ATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTT 891
Db        249 GAGTGGCCTGGCCATGTGCATGACATCAAGGCATGGGTGGCCAATACGAGTGTATTCGTG 190

Qy        892 TTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGT 951
Db        189 CTGCCCTCCTATTATAGAGAAGGTGTGCCGAGAAGTAGTCAGGAGGCCATGTCCCTGGGG 130

Qy        952 AGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAAT 1011
Db        129 AAGCCTGTGATAACGACTGACTGGGCGGGATGTCATGAAGCGGTCCTGGATGGTTTCGAT 70

Qy       1012 GGCTTTTTGATACCTC 1027
Db         69 GGATCTCTGGTACCGC 54
```

RESULT 3
CNS0039G/c

```
LOCUS       CNS0039G                 1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
```

```
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match              6.7%;  Score 79.6;  DB 10;  Length 1101;
  Best Local Similarity   17.9%;  Pred. No. 6.9e-06;
  Matches  124;  Conservative  293;  Mismatches  273;  Indels    4;  Gaps    1;

Qy        209 TTAAGGGGGTTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTTAAGGATATAATTG 268
              :|::::: :: ::| : : : :::    ::    :::::::  ::|:::: :::  | :
Db       1076 WTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWWWWWATWDTWWDKWWWWATAAKT 1017

Qy        269 CTGTTTATGAACTAAAAAAAATTCTTAAGGATATTTCCCCAGATATTGTATTTTCATATT 328
              | :|::  |:  ::| :::     :::   :|:       |:::::::: ::  :: :
Db       1016 DTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRRDGGRKRKDKKDRKDGDDDKKG 957

Qy        329 TTGTAAAGCCAGTAATATTTGGAACTATTGCTTCAAAGTTGTCAAAAGTGCCAAGGATTG 388
              ::: ||:       ::::: :  : ::| ::  |::::|  : ::: :  :|:::|:|
Db        956 GKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDKDDDGKGKDADDDTDG 897

Qy        389 TTGGAATGATTGAAGGTCTAGGTAATGCCTTCACTTATTATAAGGGAAAGCAGACCACAA 448
              |:::::::: :::::: : |:|:  :    |: : : : :|:::  :|||: : |  : :|
Db        896 TKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTWDAAADDWWADDRWDA 837

Qy        449 AAACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGCTTG 508
              :|: :::|:    : |: ::: : |       : :::::|:::::  |:::   ||  |:
Db        836 WAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRDAADDRDDAATWTTWT 777

Qy        509 ATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTA 568
              |   :|:::: : :|: :: |:: :::::::::::| |  :: ::|  : :::::  ::
Db        776 TTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTWKRRWKRRDTRWD 717

Qy        569 AAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAAG 628
              :|: ::  :::: :: :: |:  :|: | :::  :::  ::: : :  :  :| |: :
Db        716 DADADDTARDDRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDAWWADAAWWTTTDTDT 657

Qy        629 AGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGA 688
              ::     ::: ||:::::|||   ::::|:: :    :: :::::|:|:: :::||:
Db        656 DDWDKRDRRRKGARRRRRTTARAAWDWWTWKAW----DWAKWDWKTRADRWDRWAADTWT 601

Qy        689 TATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTG 748
              :| : :::: :|   :|:  : :||: : ::|  |      |: |  |:  :   :|: |
Db        600 DARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTTATWTTWAARAAWWAWWAWATT 541

Qy        749 TAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAAT 808
              || :|||:     ||||     |: | |   :   : ||:| : |||:|  ||| : | |
Db        540 TATWTTTWTTTWTTTTTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWAAAAAAAWAATT 481

Qy        809 CGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGT 868
              |     : :  |  :| :|:||| :  |  : || |   : :| :: |  ||| :
Db        480 TTTTTTTTTTWAWTAAAWWTAWTWTTTTWTTTTWWAATTTTTTWTWTTWWAATTATTTTW 421

Qy        869 TAGAGAAAAGTTCTGTTTTTGTTTTACCTACATC 902
              |  : ||||    : :::|:  | | ||  | :|
Db        420 TTTWTAAAAAAAAWWWWWTWTATATKCCCCCCYC 387


RESULT 4
CNS00EVL
LOCUS       CNS00EVL               1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match              6.3%;  Score 74.8;  DB 10;  Length 1101;
  Best Local Similarity   33.8%;  Pred. No. 6.6e-05;
  Matches  198;  Conservative  101;  Mismatches  283;  Indels    4;  Gaps    1;

Qy        218 TTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATG 277
              ::|||:    |  |:| | |||    ::::||  ||:::  :  |: ||  : |:||:
Db        477 WWAAAWAAAWTTATWAATWAAAAAAWAWWWWATTTTTWWWWWTWATTWTTTWAWWTWTAWT 536

Qy        278 AACTAAAAAAAAATTCTTAAGGATATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGC 337
              ||  |||||:|    ||||  : || :    | |  || :| ::|    |  | || |
Db        537 AAAAAAAAAWATAATTTAAAWWAATAWATTAAWAATTTAWAAWWTATATTAATWTATAAA 596

Qy        338 CAGTAATATTTGGAACTATTGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGA 397
              : :| || |   ||  | |  ||      ||   |||| |   ||  |||      |:| |
Db        597 TWTWATTAATATAAAAAAATATTT----TTTWATAAAATTTTTAATAATTTAATTAWTTA 652

Qy        398 TTGAAGGTCTAGGTAATGCCTTCACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAA 457
              || ||    |:  |:::   || | | ||| :||   :||  | |  | |||: |:||:
Db        653 TTAAATAAWTWATTWWWTAATTAAATAATTTWAAATAWAAAAAAAAAAAAAAAWATWAAW 712

Qy        458 TGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGA 517
              :| |    |:  ||| :  |  : |:  |:  : ||   :  |  || :||: |
Db        713 AATWATAWATAAWTTAAAAWAATAAAAWAAWAATWAWWATAATATWWATATATAWTTWTA 772

Qy        518 TTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAAAGCTAAGG 577
              ::   ::|::: || :  |:|| |:| : : :||  || : ||:| :|  :|:    |
Db        773 WWWATWWAWWWTATAWATAWAATAWAAWAWAWATAAATAWATAWATWAAAWAWAWATAWW 832

Qy        578 TAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGA 637
              :   | ::|       |    ||:  || |:||| |   :  | |  |: :
Db        833 ATWATATAWWAATAWAWAAAAAATWTAATATWAATWATAAWAAAAAATAWAWTTWWTWTT 892

Qy        638 AAGAGAAAATTACCTTTATTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGT 697
              : ::|  ||  : :|| :  |:|  ||    |  | | | ::   || :||
Db        893 TTWWAWWATATAAAWAWATAWAAAWAAAAAAAAAAAAAATAAWAWWWTWATATTWTTATTA 952

Qy        698 TTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAG 757
              :|: |      :|  ::  |: | |    |  :   |  | : |:| : ||  ||||
Db        953 AAWTWTATWWATTWATTWWAWTWTATAWTTWTATWTATATATWWTAWTAWATATATTTAT 1012

Qy        758 GAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAAT 803
              |  :| |      :| |:|  | | | || : : |:| || :|::
Db       1013 TAAWWTATATTTTAWAAAWTAATATATWATAWWWTAWATATAWAWW 1058
```

RESULT 5
BZ562735
LOCUS BZ562735 761 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_3906.y2 pacs2-164 Pseudomonas aeruginosa genomic clone pacs2-164_3906, genomic survey sequence.
ACCESSION BZ562735
VERSION BZ562735.1 GI:27185248
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 761)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .761
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_3906"
/clone_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun library."
ORIGIN

Query Match 6.2%; Score 74.2; DB 9; Length 761;
Best Local Similarity 52.0%; Pred. No. 8.6e-05;
Matches 166; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

Qy 838 CCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCT 897
Db 215 CCTGGCCATGTGCATGACATCAAGGCATGGGTGGCCAATACGAGTGTATTCGTGCTGCCC 274

Qy 898 ACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCT 957
Db 275 TCCTATTATAGAGAAGGAGTGCCGAGAAGTAGTCAGGAGGCCATGTGCTTGGGGAAGCCT 334

Qy 958 GTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTT 1017
Db 335 GTGATAACGACTGACTGGACGGGATGTCGTGAAACGGTACTGGATGGTTTCAATGGATTT 394

Qy 1018 TTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAAT 1077
Db 395 CTGGTACCGATCAAGAGTCCGCAGAGTTTAGCCGCGGCAATGTTGAAGTTCATCGAGTCG 454

Qy 1078 AAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGAT 1137
Db 455 CCTTCGCTCATCGCACGTATGGGAGAAGCCAGTAGATCGCTGGCGGAACAGCGCTATGAC 514

Qy 1138 GCTTTTGAAAAAAATAATA 1156
Db 515 GTTTTCAGGATAAACAATA 533

RESULT 6
BZ569481
LOCUS BZ569481 945 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_8392.y2 pacs2-164 Pseudomonas aeruginosa genomic clone pacs2-164_8392, genomic survey sequence.
ACCESSION BZ569481
VERSION BZ569481.1 GI:27204170
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 945)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .945
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_8392"
/clone_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun library."
ORIGIN

Query Match 5.7%; Score 67.8; DB 9; Length 945;
Best Local Similarity 47.9%; Pred. No. 0.0018;
Matches 226; Conservative 0; Mismatches 243; Indels 3; Gaps 1;

Qy 664 GCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAG 723
Db 294 GCTCGGTTGCTCAGGGACAAGGGTGTGCTGGAATTCGTCGAAGCGGCAAATATCCTTAGG 353

Qy 724 ACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTC 783
Db 354 CAGCGCAGCGTACCGGCACATTTTCAACTGGTCGGAGACCTGGATCCTGGGAACCCCACC 413

Qy 784 TCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGT 843
Db 414 AGTATCGAGCCGTCCGAACTAGAGCGTTGGCGTAGTGAAGGCACGATCGAGTGTCTTGGT 473

Qy 844 CATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCA 903
Db 474 TATCGCCAGGATATCGCGTCTGTCTTCG---CCCGTTCGCATATAGTCGTTCTGCCTTCT 530

Qy 904 TATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATA 963
Db 531 TATCGTGAAGGGCTTCCCAAGGTCCTGGTCGAGGCGTCGGCGTGCGGGCGTGCGGTTGTG 590

Qy 964 ACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATA 1023
Db 591 ACGACCGATGTACCGGGTTGCCGGGATGCAATCCAGGCTGATGTTACGGGCCTGCTCGTG 650

Qy 1024 CCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGAT 1083
Db 651 CCGGTGCGTGATTCTGCTGCTCTGGCGGATGCCATTCAGCGTCTTATTGAGTCGCCAGAG 710

Qy 1084 AAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTG 1135
Db 711 TTGCGCAAGAAGATGGGAGCTGCTGNGCGAGCGTTGGCAGAGCGTGATTTTG 762

RESULT 7
CNS0021J
LOCUS CNS0021J 1101 bp DNA linear GSS 03-JUN-1999

```
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL061936
VERSION     AL061936.1  GI:4940214
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR05N11"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.6%;  Score 66.6;  DB 10;  Length 1101;
  Best Local Similarity   38.3%;  Pred. No. 0.0032;
  Matches  208;  Conservative   57;  Mismatches  276;  Indels    2;  Gaps    1;

Qy        428 ATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTATATA 487
              | ||   |||  | |  | || ||  ||||  | |||    | | |||       | | |
Db        446 AAAAAAAAAAAAAAAAAAAAAAANAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 505

Qy        488 AGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATT 547
                | | ||    |  :|  |  :|| ||| : ||   |  |  |  |   :   |||
Db        506 NAANAAAAATATAATTTAWTTTTTTWTTAATTAWTTTTTTTTTTTTTTTTTTWTTAATT 565

Qy        548 TAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATC 607
              | |    | | :|:||  ||||   |:  |:    : ||   |   |::  |    |:
Db        566 TTAATTTTTAAWAWAAATTTAATAAAAWAWTWTTTAWTTTAATWTAAWWAAAAAAAAWT 625

Qy        608 TTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAA 667
              || |  : :|:|  | |  :          |:| | ||:|| :  :||:||| |::|  ||
Db        626 TTTAAAWAWTWTTTTTTTTWTTAATTAATAWAAAAAAWATAWAAWTTWTTTATWWATWAA 685

Qy        668 GGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAA 727
                : | |   | | |:      ||:||   ||: || ||     |   :|  | || :  :
Db        686 ATWAAATTTTAAAAAWTTTTTTAWTTTTTTTWTTTAAAAWTTTATTAWTTTTAAAAWATW 745

Qy        728 CTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCAT 787
              :|::  || || | |: |  |||: |||  |  |||   | |  ||||  ||| | ||
Db        746 AWTWWTAAATTTTTTAWATAWTAAWATTAWAAAATTTTTTAATT--TAATTTTTTTAAT 803

Qy        788 TACAAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATG 847
              || |||||||  |||   |:     || ||| || |  |   :|  ::|  |  | |
Db        804 TAAAAAAAAAAAAAAAAAAAWAAATAAAAAAATAAAAAAAAAAAWAAAWWTTTTTTTAAAT 863

Qy        848 TGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATC 907
              |    : | || || | |  | | |    |  |  :  | || ||| |  :  |  |:
Db        864 TATTTWTTATTTAAAAATTTAATAATSATTTAAWTATGTTTTTTAAGTWTAAAAAWTGK 923

Qy        908 GAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAA 967
              ||   :: :   : ::| : |  |: | : || |   ||  :: :     :   || ||
Db        924 GAATGKKKKGKTARKKKGRKGTTAARAATDGTAAGAKRATAAKKTKTKAATKKTAAAAAA 983

Qy        968 CTA 970
               :|
Db        984 DKA 986


RESULT 8
CNS0106X
LOCUS       CNS0106X                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098595
VERSION     AL098595.1  GI:5610206
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN03K20"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             5.5%;  Score 66;  DB 10;  Length 1101;
  Best Local Similarity   20.7%;  Pred. No. 0.0042;
  Matches  100;  Conservative  196;  Mismatches  185;  Indels    1;  Gaps    1;

Qy        252 TTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAAATTCTTAAGGATATTTCCCCAGA 311
              | ||:  : :||:||   | | |:| |  || : |:  :| | |    |:||::    :::
Db        607 TGTTDKAKTKATMATGTKTATRTWTAATKTAKRTADTKDTRTAATATAKATKKKARKDKD 666

Qy        312 TATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCTTCAAAGTTGT- 370
              ||:: : :::      :::    |     | : || :: ::|  :|: : |: | |: ::|
Db        667 TAKDAWKWDKATKAKAKKKAAKAKAKAAAKTATGKDARKAKAKAKAKATKRAKARAWDTA 726

Qy        371 CAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTCACTTATTATA 430
              : |   :  |::||::: : :|  :| :::: :  :::::|:    : | : : :|::
Db        727 TWTATAADAAADKGAKDKAKAKAKADARDKRRWDAKKDKRKRAKAKAAKAADDADADAKD 786

Qy        431 AGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGT 490
              |:  :|   |:|  | : : : :  :|::::   ::::|:      : ::  |:|:|:
Db        787 AKAADAKAKADAADADDGRGDGKKKRAKDRKKKKKKDKAWGDKKKAKKDKAAAKAKADAA 846
```

```
Qy        491 TAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAA 550
Db        847 GDAKAKRRRAGDKDKAKADAKAAKKAKAKDDDDAKAATKAKKATKAKKDKAKAKKKKKKK 906
Qy        551 TCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTA 610
Db        907 DKDAKAKAKAKDAKDKDDKDDDKDDKKDAKADKKKWKKDRAKDKAKKDDDKKDKAKDKDK 966
Qy        611 ATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGT 670
Db        967 DKKDADAKAWAKAKADADAKAKAKAKAKAADDDAAADAKAKAADAKAKAKAKADDAAKAK 1026
Qy        671 TATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTT 730
Db       1027 AKADAKAADADAKAKAAKRAAKAKADADAAKAKAKADDKDKDAKKDDADKAKKAKAKDAK 1086
Qy        731 AT 732
Db       1087 AK 1088
```

RESULT 9
CF934550
LOCUS CF934550 1091 bp mRNA linear EST 17-NOV-2003
DEFINITION TrEST-B10B02.g TrEST-B Hypocrea jecorina cDNA clone Tr-B10B02.g 5', mRNA sequence.
ACCESSION CF934550
VERSION CF934550.1 GI:38354379
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE 1 (bases 1 to 1091)
AUTHORS Bonaccorsi,E.D., Chambergo,F.S. and El-Dorry,H.
TITLE TrEST-B
JOURNAL Unpublished (2003)
COMMENT Contact: El-Dorry, Hamza
Department of Biochemistry
Institute of Chemistry. University of Sao Paulo
Avenida Professor Lineu Prestes, 748, Sao Paulo, SP, 05508-900, BRASIL
Tel: (55) 11-38183848
Fax: (55) 11-38183848
Email: dorry@iq.usp.br
PCR PRimers
FORWARD: Universal M13 forward primer
BACKWARD: Universal M13 reverse primer
Plate: 10 row: B column: 02
Seq primer: M13 reverse primer
POLYA=No.

FEATURES Location/Qualifiers
source 1. .1091
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM9414 (ATCC26921)"
/db_xref="taxon:51453"
/clone="Tr-B10B02.g"
/sex="Asexual"
/tissue_type="Mycelia"
/dev_stage="Mycelial culture grown from 100 mM and harvested to 80 mM glicose concentrations."
/lab_host="E. coli DH10B cells (streptomycin resistant)"
/clone_lib="TrEST-B"
/note="Vector: pUC18; cDNA cloned into SmaI site of pUC18. Average insert size: 1 kb"

ORIGIN

Query Match 5.5%; Score 65.4; DB 7; Length 1091;
Best Local Similarity 45.9%; Pred. No. 0.0056;
Matches 277; Conservative 0; Mismatches 324; Indels 2; Gaps 2;

```
Qy          1 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 60
Db        405 ATTGTATTATATATAATATATTTTTTTAAATATAATTTATTATAATAATATAATATTAAT 464
Qy         61 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 120
Db        465 ATATTTAATATTATTTAAATATTTTAATAATTATATTAATAATTTATTATTTAATAATAT 524
Qy        121 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 180
Db        525 ATGTAATATTTTATATATTGTATTAATATTAATTTAATATATATTGATTTTTAAATTTTA 584
Qy        181 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAAT-TCTCTCTTAACTCAAA 239
Db        585 TTTATAATATTATAATAATANTTATAATTANTTAATAATATTTATATATAAATATTATAT 644
Qy        240 GGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGA 299
Db        645 TTATTATAATATTATTTATTTTATAATTATATTGTAATAAATTTATTTATATATTNATTT 704
Qy        300 TATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGC 359
Db        705 TATATTATTTTTAATTATATAATATTATTTATTGAAATTAATTAATAGATAATATTTTAT 764
Qy        360 TTCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTT 419
Db        765 ATAANATGTATNTATATATTATATAATTATTAAGATAATTTTTATAATTAGGAATAAATT 824
Qy        420 CACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAAT-TC 478
Db        825 AAATTATTATAAATTTATATATTATTTTAATTATATAATAATGTATNTTTTAGATATATA 884
Qy        479 TTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATA 538
Db        885 ATATATATAAGTTTNATATATTAATTTTTAGATATTTATTAATATTATATTTTATTTATT 944
Qy        539 AAAAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTG 598
Db        945 ATTTATAATTATTATATAATATATTTTAATATGATAAAATAATATTATATAATGTAANTA 1004
Qy        599 GAT 601
Db       1005 TAT 1007
```

RESULT 10
CNS00EVL/c
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial

```
          EcoRI digestion of Drosophila DNA provided by the BDGP from the
          isogenic strain y2; cn bw sp, the same strain used for the BDGP's
          P1 and EST libraries. A more detailed description of the library
          and how to order individual BAC clones, the entire library, or
          filters for hybridization from the BACPAC Resource Center can be
          found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             5.5%;  Score 65.4;  DB 10;  Length 1101;
  Best Local Similarity   31.8%;  Pred. No. 0.0056;
  Matches  197;  Conservative  110;  Mismatches  311;  Indels    1;  Gaps    1;

Qy          4 TTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAAAAA 63
              ||  ::|  :  |:| |||| |:|   |::|:| || :| ::  ||   |||| |:   :
Db       1086 TTTTWWTATWTTAWATATATAAWTAATTWWTWTATATWTAWWWTATWATATATTAWTTTW 1027

Qy         64 ATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATCATG 123
                |   || ::    |  ||   : |:| :: |  |  :  |  |:   |: : : |
Db       1026 TAAAATATAWWTTAATAAATATATWTAWTAWWATATATAWATAWAAWTATAWAWTWWAAT 967

Qy        124 AATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAAGAT 183
              :| |:: | :       | |:||| |: ::: : ||      |  |||||    : |   :
Db        966 WAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTTTTTTWTTTWTATWTW 907

Qy        184 CTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAGGGT 243
              ||| | :: ::     |:       |       ||| : | : | |||: |       |
Db        906 TTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATATTAWAT-TTTTTWT 848

Qy        244 ATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGATATT 303
              :|   :   | |:|  :|||: :|  | : |:|  :     | : :|: |: |  :|||:
Db        847 WTATTWWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTWTWTTWTATTWTATW 788

Qy        304 TCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCTTCA 363
              |    :  :: | ::|  :  :||| |  ::|    | :: :|||:  :  |  |  ||
Db        787 TATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTTWTTTTATTWTTTT 728

Qy        364 AAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTCACT 423
              || || |  | : |    :  :|| ||    |  || :   | :|  | ||    | | :
Db        727 AAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTWTATTTWAAATTATTTAATTAWW 668

Qy        424 TATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTA 483
              :| |: :      |  | :  |  |||     ||  ||    |  | | | ||| |||||
Db        667 WAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAAAAATATTTTTTTA 608

Qy        484 TATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAAAAA 543
              ||| | |:|   |||   ||    |    :||: |   |:|||| : || ::  ||||  |
Db        607 TATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTATTWWTTTAAATTA 548

Qy        544 GATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTG 603
              :|||  |       :|:: :|  :||   ::  :|| | | :::  |    :   : |
Db        547 TWTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWTWTTTTTWATTWATA 488

Qy        604 GATCTTAATGAGTTTTCAT 622
              :| |:  | : : || ::
Db        487 AWTTTWTTTWWAWATTWWK 469


RESULT 11
BZ569436
LOCUS       BZ569436                 920 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_8379.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_8379, genomic survey sequence.
ACCESSION   BZ569436
VERSION     BZ569436.1  GI:27204054
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 920)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .920
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_8379"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             5.4%;  Score 64.6;  DB 9;  Length 920;
  Best Local Similarity   47.5%;  Pred. No. 0.0081;
  Matches  224;  Conservative    0;  Mismatches  245;  Indels    3;  Gaps    1;

Qy        664 GCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAG 723
              ||  ||||  | || || || ||  |  | || ||  | ||||| |||||  || ||| |
Db        340 GCTCGGTTGCTCAGGGACAAGGGTGTGCTGGAATTCGTCGAAGCGGCAAATATCCTTAGG 399

Qy        724 ACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTC 783
                      |    | | |||  | |  | ||||   | ||   |   || ||   |
Db        400 CAGCGCAGCGTACCGGCACATTTTCAACTGGTCGGAGACCTGGATCCTGGGAACCCCACC 459

Qy        784 TCATTACAAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGT 843
                  |  |      ||| | ||        | |  |||         |    | || ||||
Db        460 AGTATCGAGCCGTCCGAACTAGAGCGTTGGCGTAGTGAAGGCACGATCGAGTGTCTTGGT 519

Qy        844 CATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCA 903
                ||    |  || |       |   || |      ||||   | | ||  | |   | ||
Db        520 TATCGCCAGGATATCGCGTCTGTCTTCG---CCCGTTCGCATATAGTCGTTCTGCCTTCT 576

Qy        904 TATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATA 963
              ||||| |||||  | || | |||  |     || ||   |||     ||  |  | ||  |
Db        577 TATCGTGAAGGGCTTCCCAAGGTCCTGGTCGAGGCGTCGGCGTGCGGGCGTGCGGTTGTG 636

Qy        964 ACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATA 1023
              || ||  ||||||| || ||  |||||  |||  | | ||  |  |  ||| |  |  |
Db        637 ACGACCGATGTACCGGGTTGCCGGGATGCAATCCAGGCTGATGNTACGGGCCTGCTCGTG 696

Qy       1024 CCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGAT 1083
              ||      |||   |  |   |||| ||    ||  |   | |||||||      |||
Db        697 CCGGTGCGTGATTCTGCTGCTCTGGGCGATGCCATTCAGCGTCTTATTGAGTCGCCAGAG 756

Qy       1084 AAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTG 1135
                   | |||||   ||||||  |   ||| ||| |      | ||||
Db        757 TTGCGCAAGAAGATGGGAGCTGCTGGGCGAGCGTTGGCAAAGCGTGATTTTG 808


RESULT 12
CC264939/c
LOCUS       CC264939                 1272 bp    DNA     linear   GSS 13-MAY-2003
DEFINITION  CH261-19L17_Sp6.1 CH261 Gallus gallus genomic clone CH261-19L17,
            genomic survey sequence.
ACCESSION   CC264939
```

```
VERSION     CC264939.1  GI:30612358
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 34
            High quality sequence stop: 261.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-19L17"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN

  Query Match             5.3%;  Score 63.6;  DB 9;  Length 1272;
  Best Local Similarity   43.2%;  Pred. No. 0.013;
  Matches  378;  Conservative    0;  Mismatches  489;  Indels    7;  Gaps    2;

Qy         31 TATTATTTACTTTTAGAATGTGATATGAAAAAAATAGTGATCATAGGCAATGTAGCGTCA 90
Db       1211 TATAATATAAATATTAAATATGATATAAAAATAATTAAAATANAATAAAATATAAATATA 1152

Qy         91 ATGATGTTAAGGTTCAGGAAAGAATTAATCATGAATTTAGTGAGGCAAGGTGATAATGTA 150
Db       1151 ATATATATTAAATNAAAATAATGAAATATAAAAAAAAAAAAATTATATTATAATTAATTAA 1092

Qy        151 TATTGTCTAGCAAATGATTTTTCCACTGAAGATCTTAAAGTACTTTCGTCATGGGGCGTT 210
Db       1091 TAATAAATTTTTAATTTAATATATATAATAAATATAATATATATAAAATATAATATATTT 1032

Qy        211 AAGGGGGTTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTTAAGGATATAATTGCT 270
Db       1031 TAATATAATAATAAATNATAAATAAAATAATATATATAATAAAAATAAAATATAAATTTT 972

Qy        271 GTTTATGAACTAAAAAAAATTCTTAAGGATATTTCCCCAGATATTGTATTTTCATATTTT 330
Db        971 TTATTAAAAAAAATTTATATTTTAAATTTTATAAATATATAATTTTAATAAAATATATAT 912

Qy        331 GTAAAGCCAGTAATATTTGGAACTATTGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGTT 390
Db        911 ATAAATTTAATAAAATATATTAATATAATTAAAATATTATTTTATTTAAAAAAAAAATTA 852

Qy        391 GGAATGATTGAAGGTCTAGGTAATGCCTTCACTTATTATAAGGGAAAGCAGACCACAAAA 450
Db        851 ATAAAATATATATAATTATAATATTATTAAAATTTATATATAATTAAAAGAAATAAATAA 792

Qy        451 ACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGCTTGAT 510
Db        791 AAAATAATAAATAAATATATTTAAATTATATAATATTATATATTAAATAAAATATTAATA 732

Qy        511 GATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAAA 570
Db        731 AAATTAAATATTAAAAATAATAAATTTAATAAAAATATAATAATAATAAAAAATTTTAAT 672

Qy        571 GCTAAGGTAACAGTGTT------AGGTGGGATTGGATTGGATCTTAATGAGTTTTCATAT 624
Db        671 AAATAATTAATAATATTTTAATAATATATATTTAAATTAAATAAATAATTAAAAATAT 612

Qy        625 AAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAA 684
Db        611 TTTATAATATATATAATAAAAATTAAATAAAATAAAATANAATTATTATTATAATAAAAT 552

Qy        685 GGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAA 744
Db        551 ATTATATTAAAATTANTATTATTTATAAATAATTAAAATTAATATATTATAATAAAAAAA 492

Qy        745 TTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTT-CTCATTACAAAAAAATGAAAT 803
Db        491 AAATTTATAATTTAATTATAAAAAATAAATAATAATTTATTTAATAAAAATAATANAAAT 432

Qy        804 TGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGA 863
Db        431 TTATTTATAAATAATTNTAATAAAAATTACATATAATAATTATATTAATAAATTATAATA 372

Qy        864 TTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCT 897
Db        371 TTTTTTTTAAAAAAATAATAATTTTAATNTTNNT 338


RESULT 13
BZ569276
LOCUS       BZ569276                 903 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_8326.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_8326, genomic survey sequence.
ACCESSION   BZ569276
VERSION     BZ569276.1  GI:27203626
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 903)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .903
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_8326"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             5.3%;  Score 62.8;  DB 9;  Length 903;
  Best Local Similarity   52.7%;  Pred. No. 0.019;
  Matches  136;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy        878 GTTCTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAG 937
Db        240 GTTCGCATATAGTCGTTCTGCCTTCTTATCGTGAAGGGCTTCCCAAGGTCCTGGTCGAGG 299
```

```
Qy        938 CTATGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAA 997
              |   |||    ||  |  | ||  | || ||  ||||||| || ||  |||||  |||
Db        300 CGTCGGCGTGCGGGCGTGCGGTTGTGACGACCGATGTACCGGGTTGCCGGGATGCAATCC 359

Qy        998 ATGATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAA 1057
              | | ||  || |  ||| |  |  | ||      |||   |  |   ||||| ||    |
Db        360 AGGCTGATGTTACGGGCCTGCTCGTGCCGGTGCGTGATTCTGCTGCTCTGGCGGATGCCA 419

Qy       1058 TGAAATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGT 1117
              |   | ||||||||     |||          | |||||   ||||||  |   ||
Db        420 TTCAGCGTCTTATTGAGTCGCCAGAGTTGCGCAAGAAGATGGGAGCTGCTGGGCGAGCGT 479

Qy       1118 TTGCAGAAAAAAACTTTG 1135
              | |||||     | ||||
Db        480 TGGCAGAGCGTGATTTTG 497


RESULT 14
BZ567774/c
LOCUS       BZ567774               1150 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_7205.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_7205, genomic survey sequence.
ACCESSION   BZ567774
VERSION     BZ567774.1  GI:27199602
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1150)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1150
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_7205"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             5.3%;  Score 62.8;  DB 9;  Length 1150;
  Best Local Similarity   52.7%;  Pred. No. 0.019;
  Matches  136;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy        878 GTTCTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAG 937
              ||||   | | ||  | |   | || ||||| |||||  | || | |||  |    || |
Db        650 GTTCGCATATAGTCGTTCTGCCTTCTTATCGTGAAGGGCTTCCCAAGGTCCTGGTCGAGG 591

Qy        938 CTATGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAA 997
              |   |||    ||  |  | ||  | || ||  ||||||| || ||  |||||  |||
Db        590 CGTCGGCGTGCGGGCGTGCGGTTGTGACGACCGATGTACCGGGTTGCCGGGATGCAATCC 531

Qy        998 ATGATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAA 1057
              | | ||  || |  ||| |  |  | ||      |||   |  |   ||||| ||    |
Db        530 AGGCTGATGTTACGGGCCTGCTCGTGCCGGTGCGTGATTCTGCTGCTCTGGCGGATGCCA 471

Qy       1058 TGAAATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGT 1117
              |   | ||||||||     |||          | |||||   ||||||  |   ||
Db        470 TTCAGCGTCTTATTGAGTCGCCAGAGTTGCGCAAGAAGATGGGAGCTGCTGGGCGAGCGT 411

Qy       1118 TTGCAGAAAAAAACTTTG 1135
              | |||||     | ||||
Db        410 TGGCAGAGCGTGATTTTG 393


RESULT 15
CL021844
LOCUS       CL021844               1170 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-8P15_RM1.1 CH216 Xenopus tropicalis genomic clone CH216-8P15,
            genomic survey sequence.
ACCESSION   CL021844
VERSION     CL021844.1  GI:40463657
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1170)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM1 TACGACTCACTATAGGGAGA
            Class: BAC ends
            High quality sequence start: 2
            High quality sequence stop: 339.
FEATURES             Location/Qualifiers
     source          1. .1170
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-8P15"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             5.2%;  Score 62.4;  DB 10;  Length 1170;
  Best Local Similarity   37.2%;  Pred. No. 0.023;
  Matches  302;  Conservative    0;  Mismatches  507;  Indels    3;  Gaps    2;

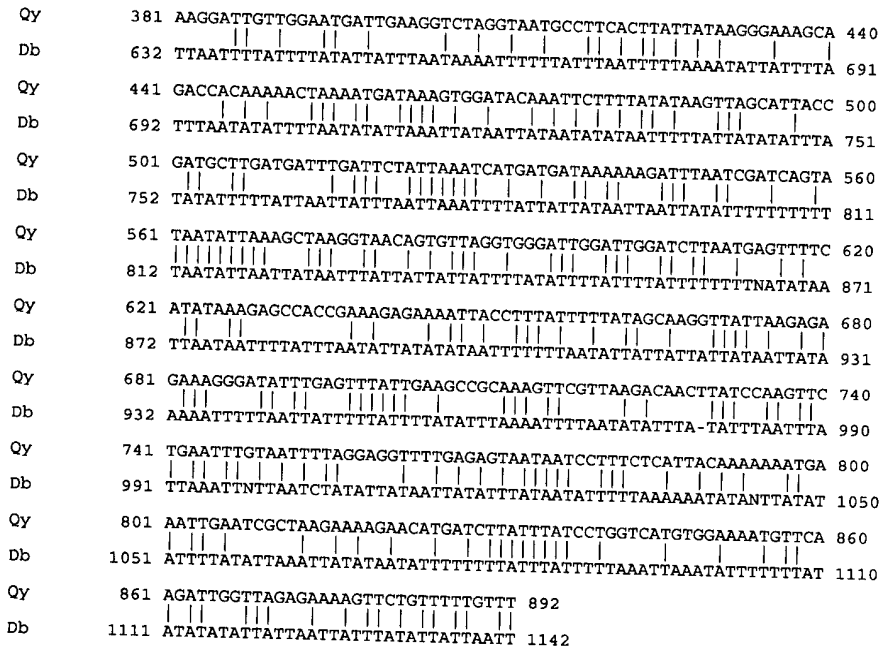
Qy         83 TAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATCATGAATTTAGTGAGGCAAGGTG 142
              ||  ||  | ||   |     ||
Db        332 TAAGGTTGANGANANTNNCCCTCCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 391

Qy        143 ATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAAGATCTTAAAGTACTTTCGTCAT 202
                  |                                                    |
Db        392 NNNNNANANNNNNNNNNNNNNNNNNNNNTNNNNNNNNNNNNNNNNNNNNNNNNANNNN 451

Qy        203 GGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTTAAGGATA 262
                                                 ||                 |  |||  | |
Db        452 NNNNNNNNNNNNNNNNNNNNNNNTNNNNAANNNNNNNNNNNNNNNTNTAATAANTAAA 511

Qy        263 TAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGATATTTCCCCAGATATTGTATTTT 322
               ||    | |  |  |    |||| || | | ||  || |       | |  |   ||||
Db        512 AAAAAATTTTAAAAAATAANAAAATAAATATAAATAATTTAATTATAAAATTATAATTTA 571

Qy        323 CATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCTTCAAAGT--TGTCAAAAGTGCC 380
               || | |    |   | |  || ||   |  |||   | ||| |  |   |||| |
Db        572 TTTAATATAATTAATAATTTTTAATTATTAAAATTATATTAAATTAATAATAAAATTATA 631
```



Search completed: May 11, 2006, 05:00:46
Job time : 3827.15 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 20:34:42 ; Search time 166.546 Seconds
(without alignments)
12743.667 Million cell updates/sec

Title: US-09-423-093B-1_COPY_739_1932
Perfect score: 1194
Sequence: 1 atgttactatgctgcataca.....aatcaaataatgatttttga 1194

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	170	14.2	1173	3	US-09-328-352-790	Sequence 790, App
	2	69	5.8	1143	3	US-09-489-039A-982	Sequence 982, App
	3	65.4	5.5	1664976	3	US-08-916-421B-1	Sequence 1, Appli
	4	65.4	5.5	1664976	3	US-09-692-570-1	Sequence 1, Appli
	5	58.4	4.9	1134	3	US-09-543-681A-3301	Sequence 3301, Ap
	6	54.4	4.6	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	7	54	4.5	50000	3	US-09-662-254B-25	Sequence 25, Appl
	8	53.8	4.5	1191	3	US-09-107-532A-1085	Sequence 1085, Ap
c	9	53.8	4.5	1683	3	US-09-809-665A-35	Sequence 35, Appl
	10	51.6	4.3	1524	3	US-09-134-001C-1141	Sequence 1141, Ap
	11	51.2	4.3	4604	3	US-09-830-807-22	Sequence 22, Appl
	12	50.2	4.2	963	4	US-09-605-703B-2155	Sequence 2155, Ap
	13	49.8	4.2	1719	3	US-09-662-254B-11	Sequence 11, Appl
c	14	49.8	4.2	50000	3	US-09-662-254B-23	Sequence 23, Appl
	15	48.8	4.1	50383	3	US-09-949-016-17600	Sequence 17600, A
c	16	48	4.0	1055	3	US-09-806-708B-23	Sequence 23, Appl
c	17	48	4.0	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	18	48	4.0	32392	3	US-09-662-254B-27	Sequence 27, Appl
	19	48	4.0	50000	3	US-09-662-254B-23	Sequence 23, Appl
c	20	46.8	3.9	1664976	3	US-08-916-421B-1	Sequence 1, Appli
c	21	46.8	3.9	1664976	3	US-09-692-570-1	Sequence 1, Appli
c	22	46.6	3.9	50000	3	US-09-662-254B-26	Sequence 26, Appl
c	23	46.2	3.9	564	3	US-08-956-171E-490	Sequence 490, App
c	24	46.2	3.9	564	3	US-08-781-986A-490	Sequence 490, App
	25	46.2	3.9	50000	3	US-09-662-254B-24	Sequence 24, Appl
c	26	45.8	3.8	6474	3	US-08-961-527-155	Sequence 155, App
	27	45.4	3.8	1182	3	US-09-134-001C-2757	Sequence 2757, Ap
	28	44.8	3.8	2848	3	US-08-936-165A-197	Sequence 197, App
c	29	44.6	3.7	2980	3	US-09-270-767-27941	Sequence 27941, A
c	30	44.6	3.7	3977	3	US-09-270-767-12208	Sequence 12208, A
	31	44.6	3.7	14602	2	US-08-597-236-1	Sequence 1, Appli
	32	44.6	3.7	14602	2	US-08-746-682A-1	Sequence 1, Appli
	33	44.6	3.7	640681	3	US-09-790-988-1	Sequence 1, Appli
c	34	44.4	3.7	18773	3	US-09-949-016-14164	Sequence 14164, A
	35	44.2	3.7	525	3	US-09-583-110-2189	Sequence 2189, Ap
	36	44.2	3.7	660	3	US-09-107-433-1123	Sequence 1123, Ap
c	37	44.2	3.7	7218	2	US-08-232-463-14	Sequence 14, Appl
c	38	44.2	3.7	50000	3	US-09-662-254B-24	Sequence 24, Appl
c	39	44.2	3.7	1830121	3	US-09-557-884-1	Sequence 1, Appli
c	40	44.2	3.7	1830121	3	US-09-643-990A-1	Sequence 1, Appli
c	41	44.2	3.7	1830121	3	US-10-158-865-1	Sequence 1, Appli
	42	44	3.7	1186	2	US-08-731-722-5	Sequence 5, Appli
c	43	43.8	3.7	832	3	US-09-621-976-2813	Sequence 2813, Ap
	44	43.8	3.7	32392	3	US-09-662-254B-27	Sequence 27, Appl
	45	43.8	3.7	50000	3	US-09-662-254B-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-09-328-352-790
; Sequence 790, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 790
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-790

Query Match 14.2%; Score 170; DB 3; Length 1173;
Best Local Similarity 52.3%; Pred. No. 1.6e-28;
Matches 492; Conservative 0; Mismatches 415; Indels 33; Gaps 4;

```
Qy      237 AAAGGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAA 296
            || ||| | |||||| ||  |||||||  |  | | |||| |  | ||||||  ||
Db      216 AACGGGAACTAATCCAATTGCAGATATAAAGACACTATTTGAAATCTATAAAATTACAAA 275

Qy      297 GGATATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTAT 356
            | | |||   |||||   ||| || ||||||    | ||||| || || | ||| ||  |
Db      276 GAAAATTAAGCCAGACTATGTCTTGTCATATACGATTAAGCCTGTGATTTATGGCACGCT 335

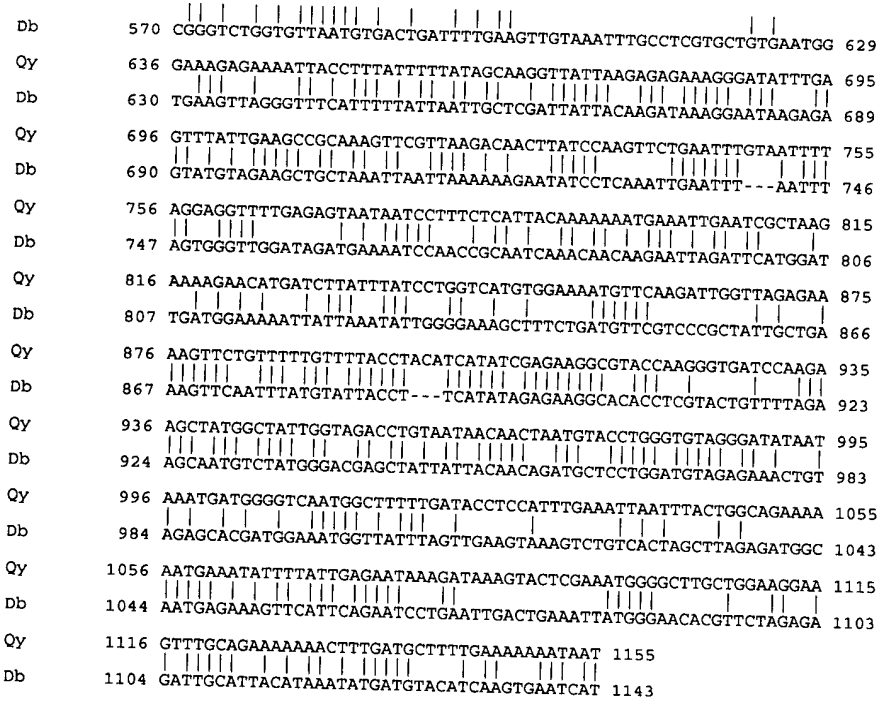
Qy      357 TGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGC 416
            |||  |  | ||| ||||||| ||||    | ||  | | ||   |||  | ||  |||
Db      336 TGCCGCTTATTTGGCAAAAGTACCAAATCGTTTTTCACTCATCACAGGATTGGGATATG- 394

Qy      417 CTTCACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAAT 476
                ||| | | || | | |    || |  |||  |  | | | |   ||   |
Db      395 -----CTTTTCAGAATGTAGACAGTACAAGTAAAAGAAGCACGTTTCAAAAAATCGTACA 449

Qy      477 TCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGA 536
            |    | |||||   ||||||| |     |  | || |   ||||| ||||||
Db      450 TTGGCTTTATAAACAAGCATTAATAAGATCATCCAAAGTTTTTTTTCAAAATCCTGATGA 509

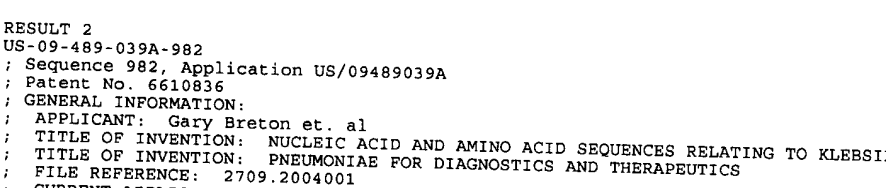
Qy      537 TAAAAAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGAT 596
                  || |   |     ||    | |    ||||||  |||   ||| ||  |
Db      510 CCTTGCTTTATTCAAACAATTAAATTTATTAAGTAATAAGGTGCCAGCCGTAGTTGTCAA 569

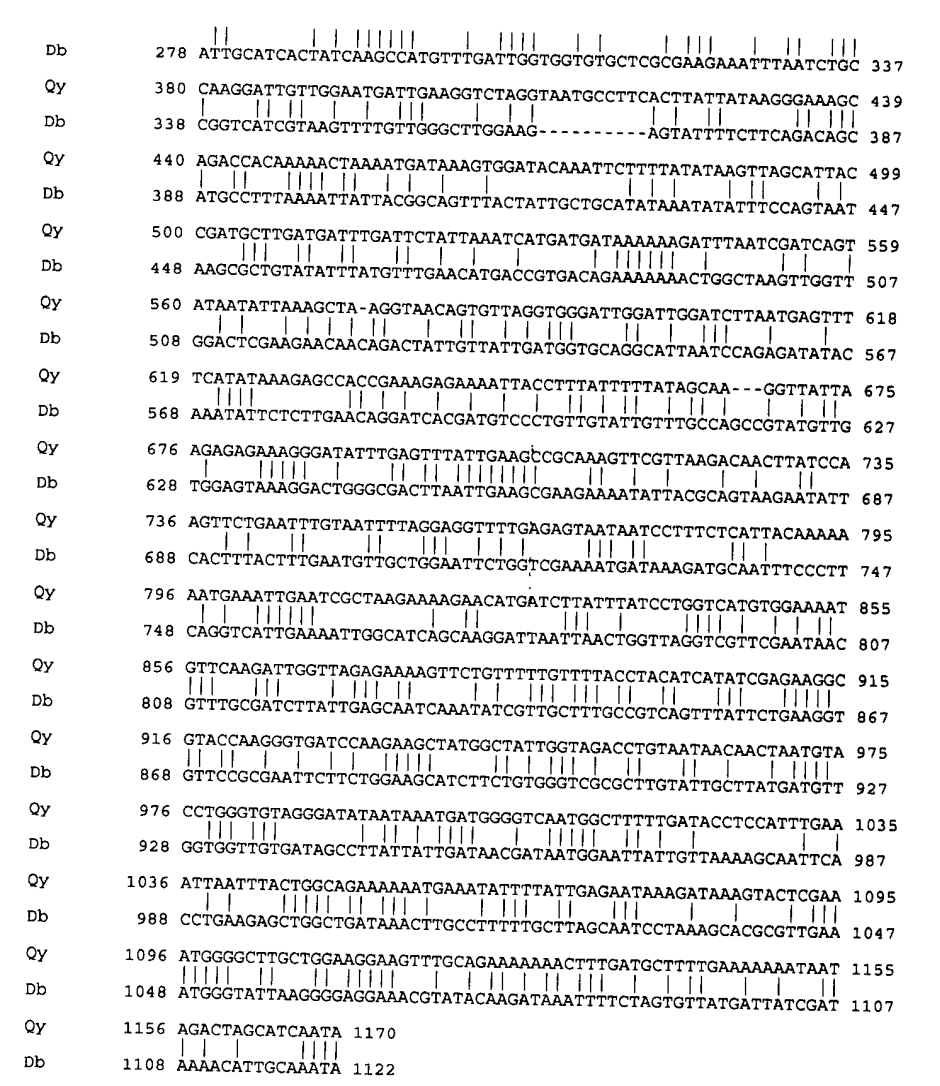
Qy      597 TGGATTGGATCTTAATGAGTTTTCATATAAA--------------------GAGCCACC 635
```



RESULT 2
US-09-489-039A-982
; Sequence 982, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 982
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-982

Query Match 5.8%; Score 69; DB 3; Length 1143;
Best Local Similarity 44.2%; Pred. No. 4.3e-06;
Matches 431; Conservative 0; Mismatches 530; Indels 14; Gaps 3;





RESULT 3
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428

; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 5.5%; Score 65.4; DB 3; Length 1664976;
Best Local Similarity 52.4%; Pred. No. 0.00014;
Matches 144; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
Qy        854 ATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAAG 913
              |||    || ||  ||   ||  |  | |  |||| |  ||  | | |||    | ||||
Db     355625 ATGAAGTAGCTTCATTTATGAGGAAGTGTAGTTTTTTAGTAGTTCCTTCAAGAAGTGAAG 355684

Qy        914 GCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAATG 973
              |   |   || ||||          ||||  ||||||     ||  |||||||||| | || |  |
Db     355685 GTTTTGGAATGGTGGCTGTTGAAGGAATGGCTTGCTCTAAGCCTGTAATAGCCACAAGGG 355744

Qy        974 TACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTTG 1033
              |    ||||   |||| ||  | | ||||||    || ||  | |||          |  |
Db     355745 TTGGGGGGTTGGGGGAGATTGTTATTGATGGATATAACGGACTATTGGCTGAGAAAAATA 355804

Qy       1034 AAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAAGTACTCG 1093
              |    |||    |   |||||||||     | || ||  | ||| ||||   |  |
Db     355805 ACCCAAATGATTTAAAAGAAAAAATTCTGGAGTTAATAAATAATGAAGAACTAAGAAAAA 355864

Qy       1094 AAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAA 1128
              |||||     |||||   | ||| || |||||
Db     355865 CTTTGGGGGAAAATGGAAAAGAATTTTCAAAAAAA 355899
```

RESULT 4
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6797466
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; CURRENT FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)

; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g

Query Match 5.5%; Score 65.4; DB 3; Length 1664976;
Best Local Similarity 52.4%; Pred. No. 0.00014;
Matches 144; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
Qy        854 ATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAAG 913
              |||    || ||  ||   ||  |  | |  |||| |  ||  | | |||    | ||||
Db     355625 ATGAAGTAGCTTCATTTATGAGGAAGTGTAGTTTTTTAGTAGTTCCTTCAAGAAGTGAAG 355684

Qy        914 GCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAATG 973
              |  |   || ||||      |||| ||||||     ||  |||||||||| | || |  |
Db     355685 GTTTTGGAATGGTGGCTGTTGAAGGAATGGCTTGCTCTAAGCCTGTAATAGCCACAAGGG 355744

Qy        974 TACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTTG 1033
              |    ||||   |||| ||  | | ||||||    || ||  | |||         |  |
Db     355745 TTGGGGGGTTGGGGGAGATTGTTATTGATGGATATAACGGACTATTGGCTGAGAAAAATA 355804

Qy       1034 AAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAAGTACTCG 1093
              |   |||    |   |||||||||     | || ||  | ||| ||||   |  |
Db     355805 ACCCAAATGATTTAAAAGAAAAAATTCTGGAGTTAATAAATAATGAAGAACTAAGAAAAA 355864

Qy       1094 AAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAA 1128
                |||||     |||||   | ||| || |||||
Db     355865 CTTTGGGGGAAAATGGAAAAGAATTTTCAAAAAAA 355899
```

RESULT 5
US-09-543-681A-3301
; Sequence 3301, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 3301
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-3301

Query Match 4.9%; Score 58.4; DB 3; Length 1134;
Best Local Similarity 47.3%; Pred. No. 0.00097;
Matches 176; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

```
Qy        803 TTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAG 862
Db        728 TTAAACTTGCAGGATTAGAAGAGAGTGTCTTTTTCTTAGGCAACCGCAATGATGTGCCGG 787

Qy        863 ATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAA 922
Db        788 ATTGCTTAAATGCAATGGATTTATTTGCCTTACCCTCTTTTGGTAACGAAGGGGTGCCAC 847

Qy        923 GGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGT 982
Db        848 AAGGTATTATGCAAGCGATGGCTTGCGGATTACCGGTGGTATCAACCACAGTAGGTGCAA 907

Qy        983 GTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATT 1042
Db        908 TTAGTGAAGCGGTGATTGATGGTAAAACCGGCTTTACATTGGCACCTAAAGTACAAGAAA 967

Qy       1043 TACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGC 1102
Db        968 CCCTAATTAATTACCTTGCTAAATTAATGGCCTCTGATGAGTTACGTCAACAAATGGGGC 1027

Qy       1103 TTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCTTTTGAAAAAAATAATAGACTAG 1162
Db       1028 AAGCGTCATTAGCCCATGCCAAAGCTCAGTTTGGTTTAGATAACATGCTAGATAAAATGG 1087

Qy       1163 CATCAATAATAA 1174
Db       1088 AAAAAATATTTA 1099
```

RESULT 6
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 4.6%; Score 54.4; DB 3; Length 1141;
Best Local Similarity 10.2%; Pred. No. 0.0075;
Matches 83; Conservative 297; Mismatches 433; Indels 0; Gaps 0;

```
Qy        364 AAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTCACT 423
Db         61 WARMYCKYRRWYNNKSRWWKGWYKKKWYBCANNTSBRYHARRWKDMKTAYBMTMTNKWGK 120

Qy        424 TATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTA 483
Db        121 TGWRHRYWRWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNATGWDDDTK 180

Qy        484 TATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAAAAA 543
Db        181 YHMWNNNGCBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRKVRR 240
```

```
Qy        544 GATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTG 603
Db        241 WVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRAKYWGWNRA 300

Qy        604 GATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATA 663
Db        301 BVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNWTYKKA 360

Qy        664 GCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAG 723
Db        361 RHBARWDWVWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMDMD 420

Qy        724 ACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTC 783
Db        421 WBGTYNNNNNGGRTYYGWTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKTYWN 480

Qy        784 TCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGT 843
Db        481 NCYWKSMTNGKSHRBAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBVCSKWWNNY 540

Qy        844 CATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCA 903
Db        541 AAWYTKSSWNYTSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWSHTW 600

Qy        904 TATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATA 963
Db        601 BHBRAGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNNNNNATCARD 660

Qy        964 ACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATA 1023
Db        661 DYYAASRWYAMANAKWYYYKBAANNAYYTHANNWWGCWNNATDTRRTMWKNNNNNNAGTW 720

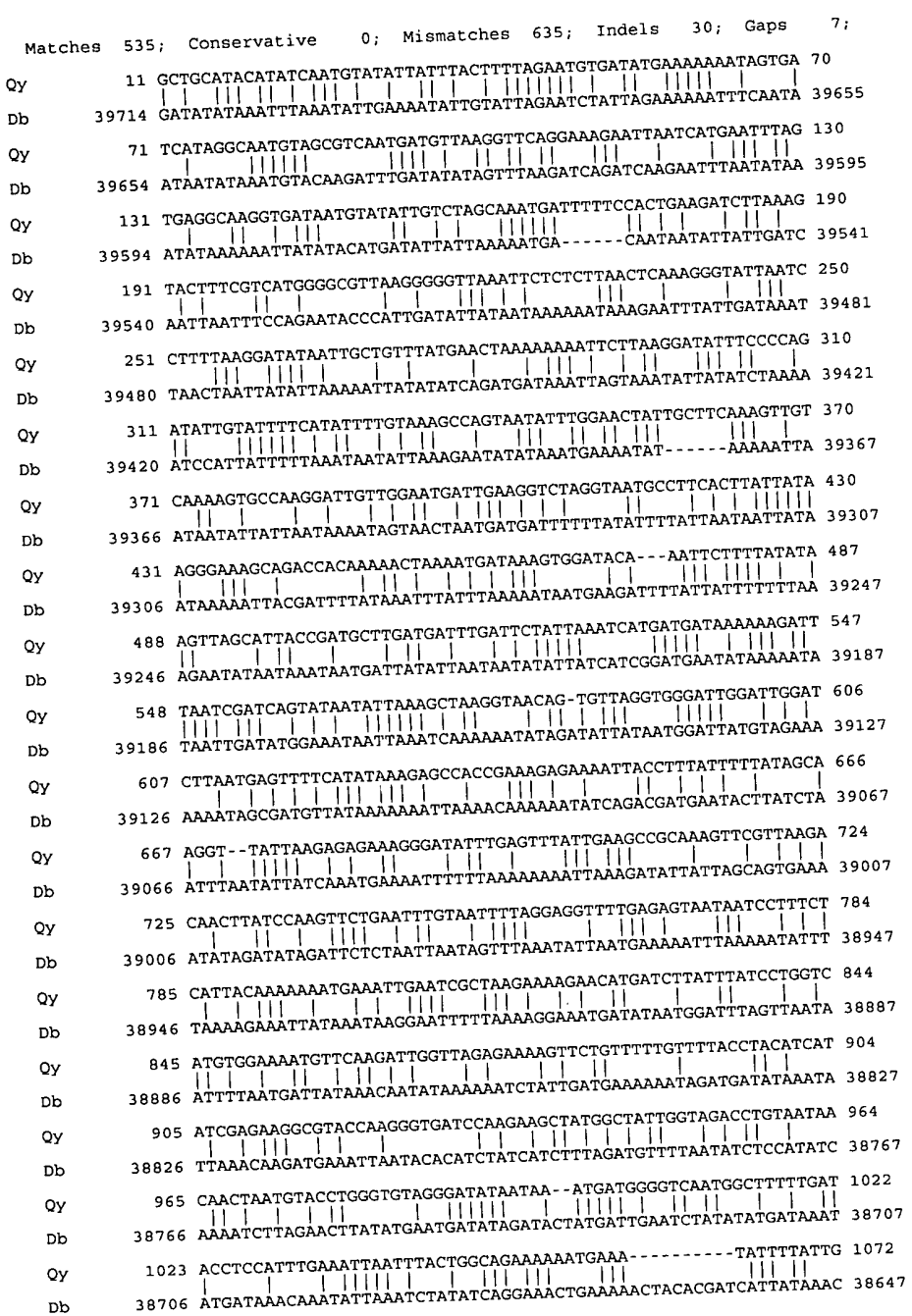
Qy       1024 CCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGAT 1083
Db        721 KNNNNNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADAAABTTDKRNNGAYTKY 780

Qy       1084 AAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCTTTT 1143
Db        781 TTTNNNNTYRGVVTNTAARDGWANNNNNNNNNNNNNNNGWSDMWVTWWAYANYGTNNNNNN 840

Qy       1144 GAAAAAAATAATAGACTAGCATCAATAATAAAA 1176
Db        841 NNNNAYAWWTNKWYYTTDDRWRBAYTNNNNNNR 873
```

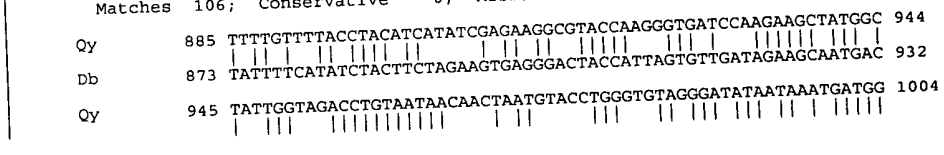
RESULT 7
US-09-662-254B-25/c
; Sequence 25, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-25

Query Match 4.5%; Score 54; DB 3; Length 50000;
Best Local Similarity 44.6%; Pred. No. 0.021;



RESULT 8
US-09-107-532A-1085
; Sequence 1085, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1085:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1191 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...1191
; SEQUENCE DESCRIPTION: SEQ ID NO: 1085:
US-09-107-532A-1085

Query Match 4.5%; Score 53.8; DB 3; Length 1191;
Best Local Similarity 54.9%; Pred. No. 0.01;
Matches 106; Conservative 0; Mismatches 87; Indels 0; Gaps 0;



```
Db        933 TTATGGATTACCTGTAATAAGTTTTGATCATGGTGGTGCTAAGGAAATATTACAAGATGG 992
Qy       1005 GGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATA 1064
Db        993 AAATTATGGAATTTTAATTAAGAATTTTGACGTACAAGATATGAATAATGAGATAGAAAA 1052
Qy       1065 TTTTATTGAGAAT 1077
Db       1053 ATTGTCTAATAAT 1065
```

RESULT 9
US-09-809-665A-35/c
; Sequence 35, Application US/09809665A
; Patent No. 6790950
; GENERAL INFORMATION:
; APPLICANT: Lowery E., David, et al.
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
; FILE REFERENCE: 28341/00435
; CURRENT APPLICATION NUMBER: US/09/809,665A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/153,453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128,689
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 09/545,199
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Pasteurella multocida
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (325)..(1230)
; OTHER INFORMATION: lgtC
; NAME/KEY: misc_feature
; LOCATION: 219
; OTHER INFORMATION: Xaa = any or unknown amino acid
; NAME/KEY: misc_feature
; LOCATION: 226
; OTHER INFORMATION: Xaa = any or unknown amino acid
; NAME/KEY: misc_feature
; LOCATION: 269
; OTHER INFORMATION: Xaa = any or unknown amino acid
; NAME/KEY: misc_feature
; LOCATION: 270
; OTHER INFORMATION: Xaa = any or unknown amino acid
; NAME/KEY: misc_feature
; LOCATION: 274
; OTHER INFORMATION: Xaa = any or unknown amino acid
US-09-809-665A-35

Query Match 4.5%; Score 53.8; DB 3; Length 1683;
Best Local Similarity 53.4%; Pred. No. 0.011;
Matches 159; Conservative 0; Mismatches 133; Indels 6; Gaps 2;

```
Qy        792 AAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGA 851
Db       1627 AGAGAAACAAATAGAATCACTCAATCTACAAGAAGATTGTTTTCTTTTAGGAAAT---AA 1571
Qy        852 AAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGA 911
Db       1570 AGATAATCCGTATCCATTAATAAAAAATGCTAAGCTATTTTTACACACCTCTTTGAAAGA 1511
Qy        912 AGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAA 971
Db       1510 GGGGTTACCGACAGTTATCCTAGAAAGCATGGCTTGCGGTACACCTGTAATATCCATGGA 1451
Qy        972 TGTACCT---GGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCC 1028
Db       1450 CTGCCCTACCGGTCCGAAAGAAATTCTNCGAGGAGGAGAATTTGGAGGATTAGTAAATTT 1391
```

```
Qy       1029 ATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAA 1086
Db       1390 AGGTGACGAAAATGCTTTTATACAAAAAACACTCTCTTTCCTTCAAAATCAAGATGAA 1333
```

RESULT 10
US-09-134-001C-1141
; Sequence 1141, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1141
; LENGTH: 1524
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1141

Query Match 4.3%; Score 51.6; DB 3; Length 1524;
Best Local Similarity 46.5%; Pred. No. 0.033;
Matches 276; Conservative 0; Mismatches 309; Indels 9; Gaps 3;

```
Qy        506 TTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATA 565
Db        860 TTTATAAGTATATTGTTCTCACTCACCAACAAAAGCACGATATACAGCAACATTTCCATA 919
Qy        566 TTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATA 625
Db        920 TAAGTGACGACCAATTCCAACTAGTACCACATTTTATCGAATTAGATACGGAGGTAGAAC 979
Qy        626 AAGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAG 685
Db        980 AAGATTCCTCAAACAATCAAAATCGCTTTATATATATAGGCCGTTTTAGCACAGAGAAAC 1039
Qy        686 GGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAAT 745
Db       1040 AGATTGATCATATTATTAGAGCTTATCACAAGTTTTTGCAAAGTGGTTACCAAACAGAAT 1099
Qy        746 TTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTG 805
Db       1100 ---TACATTTATTTGGTAGAGATGAAGATAATCAAATTCCATTGATGAATACTTTGATTT 1156
Qy        806 AATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATT 865
Db       1157 CAGAATTAAAATTATCGGATAAAGTTAAAATATTTAAATATACCAATCAACCTTTACAAA 1216
Qy        866 GGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGG 925
Db       1217 AATTTAAGAATTCTAAAGCCTCTCTACTTACAAGTCAATATGAGGGATTTGGCTTAACAC 1276
Qy        926 TGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGT- 984
Db       1277 TTAT---GGAAAGTATAGAAATGGGGTGTCCAGTCCTATCTTATAACGTTCGTTACGGTC 1333
Qy        985 --AGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATT 1042
Db       1334 CAAGTGAAATTATTCAAAACGGGATAAATGGCTATCTCATTGAAAAAAATGATATTGATA 1393
Qy       1043 TACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAAGTACTCGAAA 1096
Db       1394 GTTTATCAAAACATATGATTAACATCATTGAGCACCCACTACAAAAAGTGAAAA 1447
```

```
RESULT 11
US-09-830-807-22
; Sequence 22, Application US/09830807
; Patent No. 6846667
; GENERAL INFORMATION:
;  APPLICANT: Crooke,Helen R.
;  APPLICANT:  Clarke, Enda E.
;  APPLICANT:  Everest, Paul H.
;  APPLICANT:  Dougan, Gordon
;  APPLICANT:  Holden, David W.
;  APPLICANT:  Shea, Jacqueline E.
;  APPLICANT:  Feldman, Robert G.
;  TITLE OF INVENTION: VIRULENCE GENES AND PROTEINS, AND THEIR USE
;  FILE REFERENCE: GJE-65
;  CURRENT APPLICATION NUMBER: US/09/830,807
;  CURRENT FILING DATE:  2001-04-30
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
;   LENGTH: 4604
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2982)..(4082)
;   NAME/KEY: CDS
;   LOCATION: (1534)..(2637)
;   NAME/KEY: CDS
;   LOCATION: (749)..(1531)
;   NAME/KEY: CDS
;   LOCATION: (6)..(746)
US-09-830-807-22

  Query Match             4.3%;  Score 51.2;  DB 3;  Length 4604;
  Best Local Similarity   48.8%;  Pred. No. 0.052;
  Matches  138;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        874 AAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAA 933
              |||  |  || | || || ||    | ||   | | || ||  | |||   |  ||   |
Db       3777 AAAGTTAATGATCTTTTTATATTACCTTCTCTTTGGGAGGGTATGCCATTAGCAATTTTA 3836

Qy        934 GAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATA 993
              |||||  ||     |||    ||    |||  |||||| | || ||   ||      ||
Db       3837 GAAGCATTGAGCTGTGGACTTCCATGTATAGTCACTAATATTCCAGGTAATAATAGCTTA 3896

Qy        994 ATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAA 1053
              ||| | |||||    ||||| | |||| |          ||   | | ||| |  |  ||
Db       3897 ATAGAAGATGGCTATAATGGTTGTTTGTTTGAAATTAGAGATTGTCAGTTATTATCTCAA 3956

Qy       1054 AAAATGAAATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGG 1113
              |||||  |  |  | | |||  ||   |||     |     |            || | |
Db       3957 AAAATCATGTCATATGTTGGTAAGCCAGAACTGATTGCACAGCAATCTACCAATGCACGA 4016

Qy       1114 AAGTTTGCAGAAAAAAACTTTGATGCTTTTGAAAAAAATAATA 1156
                 |||      |||||  | ||      || ||| ||||||||
Db       4017 TCATTTATTCTGAAAAATTATGGATTAGTTAAAAGAAATAATA 4059


RESULT 12
US-09-605-703B-2155
; Sequence 2155, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
;  APPLICANT: Pompejus, Markus
;  APPLICANT:  Kroger, Burkhard
;  APPLICANT:  Schroder, Hartwig
;  APPLICANT:  Zelder, Oskar
;  APPLICANT:  Haberhauer, Gregor
;  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
;  TITLE OF INVENTION:  PROTEINS
;  FILE REFERENCE: BGI-129CP
;  CURRENT APPLICATION NUMBER: US/09/605,703B
;  CURRENT FILING DATE:  2000-06-27
;  PRIOR APPLICATION NUMBER: 60/142,764
;  PRIOR FILING DATE: 1999-07-08
;  PRIOR APPLICATION NUMBER: 60/152,318
;  PRIOR FILING DATE: 1999-09-03
;  NUMBER OF SEQ ID NOS: 2934
; SEQ ID NO 2155
;   LENGTH: 963
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (101)..(940)
;   OTHER INFORMATION: RXA01268
US-09-605-703B-2155

  Query Match             4.2%;  Score 50.2;  DB 4;  Length 963;
  Best Local Similarity   50.6%;  Pred. No. 0.062;
  Matches  121;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy        882 TGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTAT 941
              || | || |  | |   || |  | |||||||  ||   | |    | |  ||||||
Db        634 TGATGTTTTGCTTCACCCAACTCAGCGAGAAGGTTTAGGTATGTCTTTGCTGGAAGCTCA 693

Qy        942 GGCTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGA 1001
              ||||||  ||    ||||||  | || | |  |||  ||||       |||  | |||
Db        694 GGCTATGGGAGTGCCTGTATTGACGAATGCTGTGACTGGAACAGTTGATGCAGTAACAAG 753

Qy       1002 TGGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAA 1061
              |||  |    ||||||||     |      |   || | |   ||     || ||  |
Db        754 TGGAGAAGGTGGCTTTTTTGCCGATGACGATTCTGTTGAGTCCTGGGTTTCTAAGATTGA 813

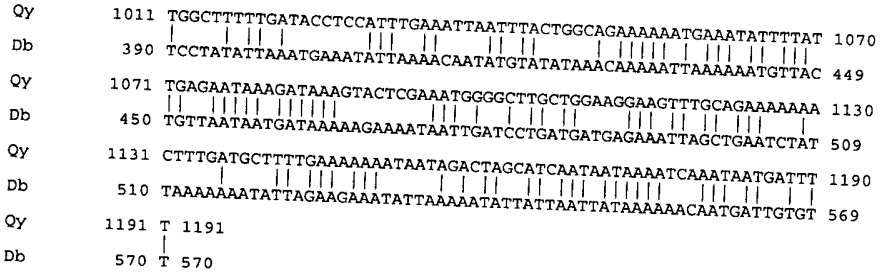
Qy       1062 ATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTG 1120
                |  ||  ||    ||    |   |  |  |   ||||| | ||||||  |  ||||||
Db        814 TTTATTAGTTTCCGATCCTAAGTTAAGAGACCGGATGGGACGTGCTGGTCGCCAGTTTG 872


RESULT 13
US-09-662-254B-11
; Sequence 11, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
;   LENGTH: 1719
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-11

  Query Match             4.2%;  Score 49.8;  DB 3;  Length 1719;
  Best Local Similarity   47.8%;  Pred. No. 0.086;
  Matches  144;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

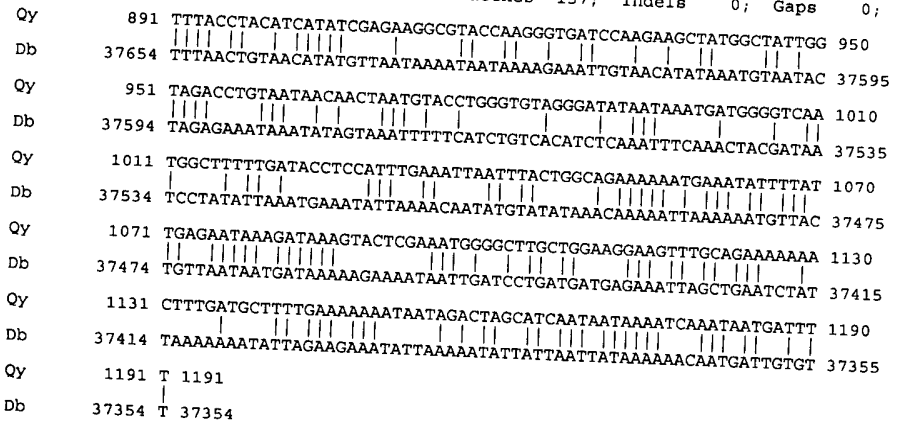
Qy        891 TTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGG 950
              |||| ||  | |||||     |      ||  ||  |  ||   | |  ||    || |
Db        270 TTTAACTGTAACATATGTTAATAAAATAATAAAAGAAATTGTAACATATAAATGTAATAC 329

Qy        951 TAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAA 1010
              ||||    |||  | |   ||| |  |       |    |  |||     |    |  ||
Db        330 TAGAGAAATAAATATAGTAAATTTTTCATCTGTCACATCTCAAATTTCAAACTACGATAA 389
```



RESULT 14
US-09-662-254B-23/c
; Sequence 23, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-23

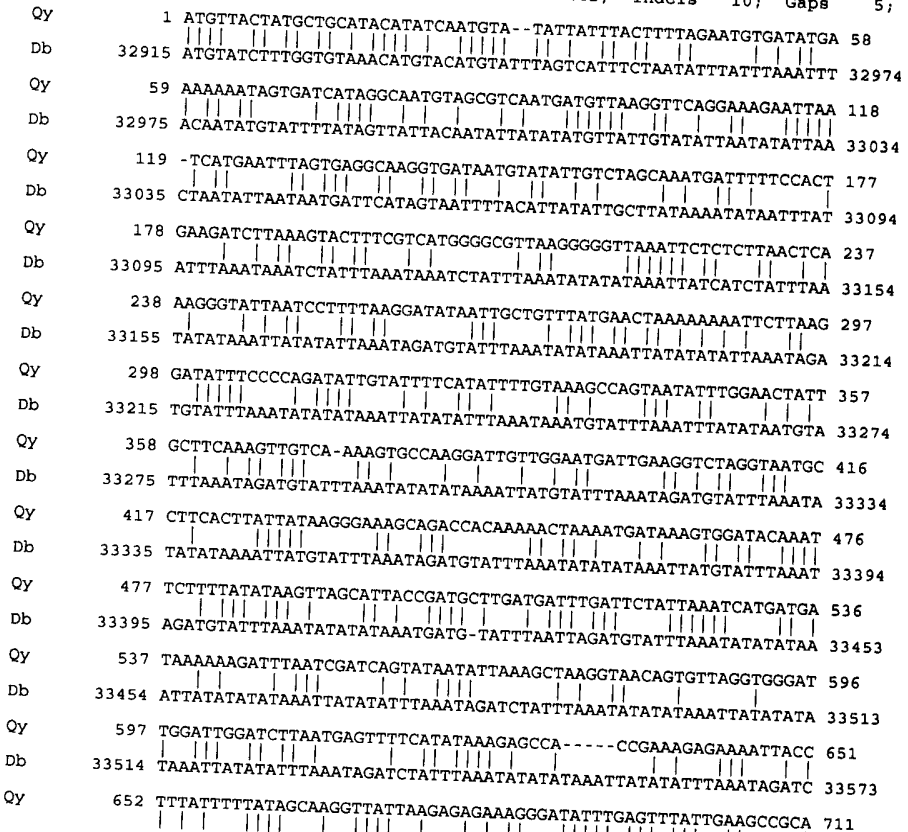
Query Match 4.2%; Score 49.8; DB 3; Length 50000;
Best Local Similarity 47.8%; Pred. No. 0.18;
Matches 144; Conservative 0; Mismatches 157; Indels 0; Gaps 0;



RESULT 15
US-09-949-016-17600

; Sequence 17600, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17600
; LENGTH: 50383
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17600

Query Match 4.1%; Score 48.8; DB 3; Length 50383;
Best Local Similarity 45.3%; Pred. No. 0.31;
Matches 374; Conservative 0; Mismatches 442; Indels 10; Gaps 5;



```
Db      33574 TATTTAAATATATATAAATTATATATTTAAATAGATCTATTTAAGTATATATAAATTATA 33633

Qy        712 AAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGT 771
              | ||   ||   | | |||   |  |    | | |||| ||  |  |   | |   | |
Db      33634 TATTTAAATAGATCTATTTAAATATTTATAAATTATGTATTTAAATAAATCTATTTAAAT 33693

Qy        772 AATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAA 817
              | | ||   || |   |  ||| ||||  | || |||   ||  ||
Db      33694 ATTTATAAATTATGTATTTAAATAAATCTATTTAAATATTTATAAA 33739


Search completed: May 11, 2006, 05:23:18
Job time : 176.546 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:01:11 ; Search time 788.075 Seconds
(without alignments)
12528.801 Million cell updates/sec

Title: US-09-423-093B-1_COPY_739_1932
Perfect score: 1194
Sequence: 1 atgttactatgctgcataca.....aatcaaataatgatttttga 1194

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1194	100.0	14516	3	US-09-423-093A-1	Sequence 1, Appli
	2	222.6	18.6	52253	9	US-10-915-740A-26	Sequence 26, Appl
c	3	222.6	18.6	2242716	9	US-10-915-740A-1068	Sequence 1068, Ap
	4	73	6.1	1134	7	US-10-282-122A-38857	Sequence 38857, A
	5	70.2	5.9	1092	7	US-10-282-122A-16045	Sequence 16045, A
	6	67.8	5.7	1080	6	US-10-369-493-24115	Sequence 24115, A
	7	65.6	5.5	1089	7	US-10-282-122A-17194	Sequence 17194, A
	8	59.8	5.0	16217	6	US-10-311-455-597	Sequence 597, App
	9	58.6	4.9	8056	8	US-10-473-126-386	Sequence 386, App
c	10	57.8	4.8	8056	8	US-10-473-126-386	Sequence 386, App
	11	57.6	4.8	5999	6	US-10-311-455-531	Sequence 531, App
	12	57.6	4.8	5999	7	US-10-257-166-41	Sequence 41, Appl
	13	57.4	4.8	15373	6	US-10-311-455-439	Sequence 439, App
	14	56.8	4.8	115218	8	US-10-278-698-255	Sequence 255, App
	15	56.8	4.8	115218	8	US-10-278-698-769	Sequence 769, App
	16	55.4	4.6	8056	8	US-10-473-126-240	Sequence 240, App
c	17	54	4.5	50000	8	US-10-706-635-25	Sequence 25, Appl
c	18	53.8	4.5	1683	3	US-09-809-665A-35	Sequence 35, Appl
c	19	53.8	4.5	1683	8	US-10-854-299-35	Sequence 35, Appl
	20	53.6	4.5	7040	5	US-10-172-086-14	Sequence 14, Appl
	21	53.6	4.5	7040	7	US-10-221-714A-162	Sequence 162, App
	22	53.6	4.5	7040	7	US-10-311-507-48	Sequence 48, Appl
	23	53.6	4.5	7040	8	US-10-480-846-14	Sequence 14, Appl
	24	53.6	4.5	7040	8	US-10-473-126-174	Sequence 174, App
	25	53.6	4.5	7040	8	US-10-473-126-320	Sequence 320, App
	26	52.8	4.4	6397	5	US-10-239-676-107	Sequence 107, App
	27	52.8	4.4	6397	6	US-10-311-455-1319	Sequence 1319, Ap
	28	52.8	4.4	6397	6	US-10-240-453-119	Sequence 119, App
	29	52.8	4.4	6397	7	US-10-221-714A-181	Sequence 181, App
	30	52.2	4.4	5992	7	US-10-221-613-168	Sequence 168, App
	31	52.2	4.4	40324	7	US-10-433-793-180	Sequence 180, App
	32	51.8	4.3	3673778	6	US-10-312-841-1	Sequence 1, Appli
	33	51.6	4.3	1524	7	US-10-724-972A-2865	Sequence 2865, Ap
	34	51.4	4.3	11691	6	US-10-311-455-2213	Sequence 2213, Ap
	35	51.2	4.3	1101	6	US-10-238-075-1319	Sequence 1319, Ap
	36	51.2	4.3	1179	6	US-10-369-493-24724	Sequence 24724, A
	37	51.2	4.3	4604	10	US-11-040-661-22	Sequence 22, Appl
c	38	51.2	4.3	6927	6	US-10-238-075-1318	Sequence 1318, Ap
	39	51.2	4.3	19087	6	US-10-311-455-766	Sequence 766, App
	40	50.6	4.2	3474	10	US-11-097-143-13003	Sequence 13003, A
	41	50.6	4.2	3677	10	US-11-097-143-12991	Sequence 12991, A
	42	50.6	4.2	5660	8	US-10-473-126-138	Sequence 138, App
	43	50.6	4.2	5666	9	US-10-486-319A-39	Sequence 39, Appl
	44	50.6	4.2	8666	6	US-10-311-455-370	Sequence 370, App
	45	50.6	4.2	8666	7	US-10-221-714A-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-09-423-093A-1
; Sequence 1, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 14516
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-423-093A-1

Query Match 100.0%; Score 1194; DB 3; Length 14516;
Best Local Similarity 100.0%; Pred. No. 1e-227;
Matches 1194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 ATGTTACTATGCTGCATACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAA 798

Qy         61 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 AAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATC 858

Qy        121 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        859 ATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAA 918

Qy        181 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        919 GATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAG 978

Qy        241 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        979 GGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGAT 1038
```

```
Qy        301 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 360
Db       1039 ATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTATTGCT 1098

Qy        361 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 420
Db       1099 TCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTC 1158

Qy        421 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 480
Db       1159 ACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAATTCTT 1218

Qy        481 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 540
Db       1219 TTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATGATAAA 1278

Qy        541 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 600
Db       1279 AAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGA 1338

Qy        601 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 660
Db       1339 TTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTT 1398

Qy        661 ATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTT 720
Db       1399 ATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTT 1458

Qy        721 AAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCT 780
Db       1459 AAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCT 1518

Qy        781 TTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCT 840
Db       1519 TTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCT 1578

Qy        841 GGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACA 900
Db       1579 GGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACA 1638

Qy        901 TCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTA 960
Db       1639 TCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTA 1698

Qy        961 ATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTG 1020
Db       1699 ATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTG 1758

Qy       1021 ATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAA 1080
Db       1759 ATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAA 1818

Qy       1081 GATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCT 1140
Db       1819 GATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCT 1878

Qy       1141 TTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1194
Db       1879 TTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1932
```

RESULT 2
US-10-915-740A-26
; Sequence 26, Application US/10915740A
; Publication No. US20050191316A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelin, Herve
; APPLICANT: Venter, J. Craig
; APPLICANT: Masignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Mora, Manrosa
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarselli, Maria
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagratia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use
; FILE REFERENCE: 002441.00090
; CURRENT APPLICATION NUMBER: US/10/915,740A
; CURRENT FILING DATE: 2004-08-11
; PRIOR APPLICATION NUMBER: 09/806,866
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: USSN 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: USSN 60/132,068
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 1068
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 26
; LENGTH: 52253
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-10-915-740A-26

Query Match 18.6%; Score 222.6; DB 9; Length 52253;
Best Local Similarity 52.4%; Pred. No. 1.1e-33;
Matches 576; Conservative 0; Mismatches 499; Indels 24; Gaps 3;

```
Qy          56 TGAAAAAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAAT 115
Db        1908 TTATGAAAATCGTTTTTATCACAACAGTCGCATCCAGCATTTACGGTTTCCGCGCCCCCG 1967

Qy         116 TAATCATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCA 175
Db        1968 TCATTAAAAAATTAATCGGCAAAAACCATCAGGTGTATGCCTTTGTATCGGAGTTTTCCG 2027

Qy         176 CTGAAGATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACT 235
Db        2028 ACAATGAATTGGATATTATCAGGGAAATGGGGGTTACACCCGTTACCTACCGTTCAAACC 2087

Qy         236 CAAAGGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTA 295
Db        2088 GCAGCGGGCTGAACCCGTTTTCGGATATAAAATCCACCTTCCTCATCTTTAAAGAACTCA 2147

Qy         296 AGGATATTTCCCCAGATATTGTATTTTCATATTTTGTAAAGCCAGTAATATTTGGAACTA 355
Db        2148 AAAAAATATCGCCGGATTTGGTTTTCCCTTATTTCGCAAAACCCGTGATTTTCGGCACTT 2207

Qy         356 TTGCTTCAAAGTTGTCAAAAGTGCCAAGGATTGTTGGAATGATTGAAGGTCTAGGTAATG 415
Db        2208 TTGCCGCAAAACTGGCAGGCGTGCCCAGAATCGTCGGGATGCTGGAAGGTTTGGGATTCG 2267

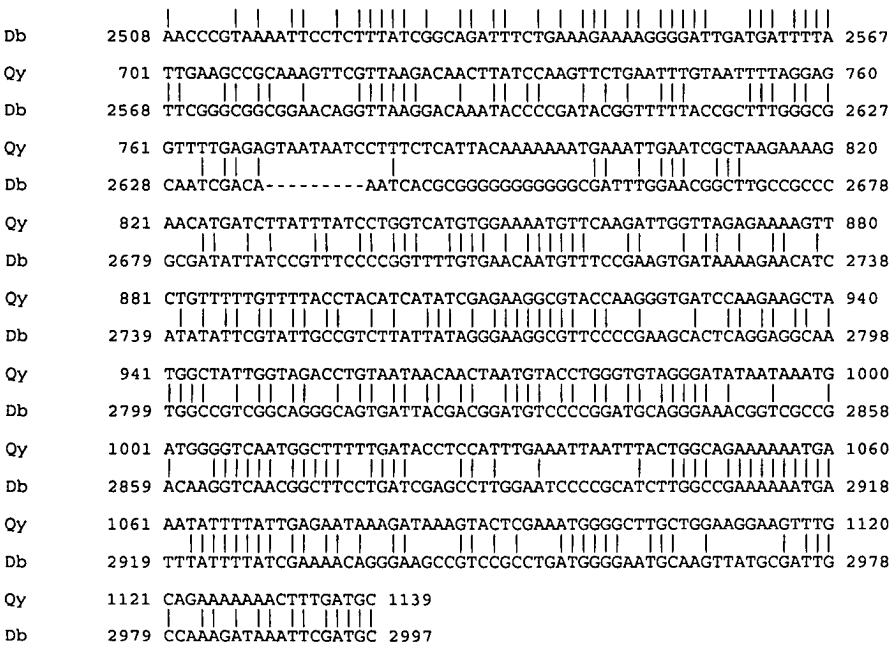
Qy         416 CCTTCACTTATTATAAGGGAAAGCAGACCACAAAAACTAAAATGATAAAGTGGATACAAA 475
Db        2268 CATTTACCCCGCAGCCGGAAGGCATACCGTTAAAAACAAAAATCATAAAGGGGATTTTGA 2327

Qy         476 TTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAATCATGATG 535
Db        2328 TTGCCTTATACCGCATTGCCCTGCCGATGTTGGAAAGCCTGATTGTATTAAACCCCGACG 2387

Qy         536 ATAAAAAAGATTTAATCGATCAGTATAATATTAA---AGCTAAGGTAACAGTGTTAGGTG 592
Db        2388 ACAAAGACGAACTGACGGACAAATACGGCATCAAAATAAAAAACATCCATATTTTGGGCG 2447

Qy         593 GGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCG----------AAAG 640
Db        2448 GAATCGGTCTGGATTTGCGGCAATATCCTTATTCCGAGGCGGATATTCCCGATGAAAAAG 2507

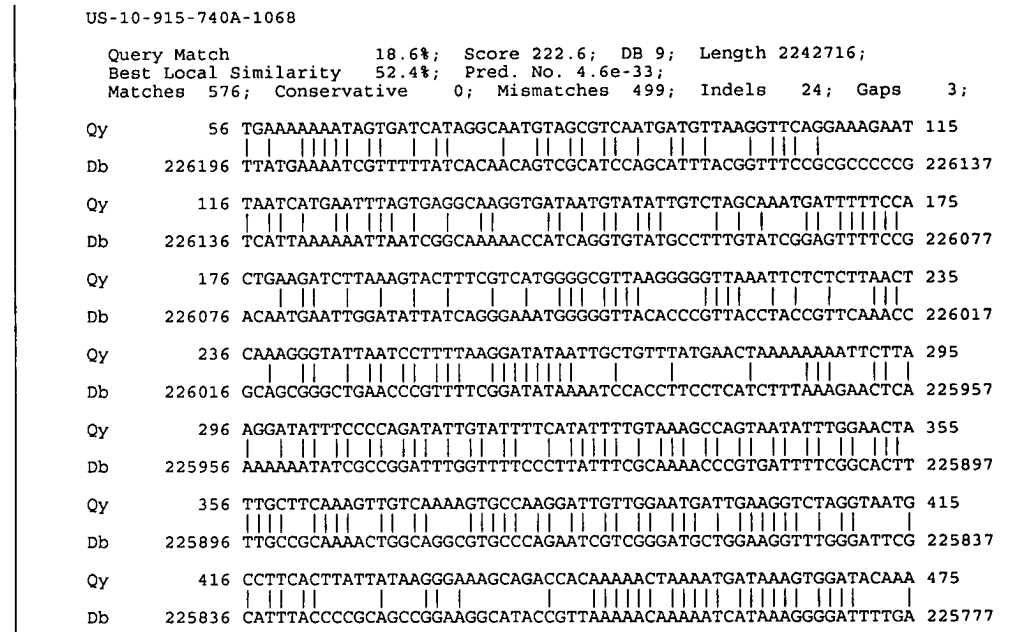
Qy         641 AGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTA 700
```



RESULT 3
US-10-915-740A-1068/c
; Sequence 1068, Application US/10915740A
; Publication No. US20050191316A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelin, Herve
; APPLICANT: Venter, J. Craig
; APPLICANT: Masignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Mora, Manrosa
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarselli, Maria
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagratia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use
; FILE REFERENCE: 002441.00090
; CURRENT APPLICATION NUMBER: US/10/915,740A
; CURRENT FILING DATE: 2004-08-11
; PRIOR APPLICATION NUMBER: 09/806,866
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: USSN 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: USSN 60/132,068
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 1068
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1068
; LENGTH: 2242716
; TYPE: DNA
; ORGANISM: Neisseria meningitidis

US-10-915-740A-1068

Query Match 18.6%; Score 222.6; DB 9; Length 2242716;
Best Local Similarity 52.4%; Pred. No. 4.6e-33;
Matches 576; Conservative 0; Mismatches 499; Indels 24; Gaps 3;



```
Qy       1061 AATATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTG 1120
             ||||||| || || | ||    || | |  ||||||  |||  |     | |||
Db     225185 TTTATTTTATCGAAAACAGGGAAGCCGTCCGCCTGATGGGGAATGCAAGTTATGCGATTG 225126

Qy       1121 CAGAAAAAAACTTTGATGC 1139
              |  || | || || |||||
Db     225125 CCAAAGATAAATTCGATGC 225107
```

RESULT 4
US-10-282-122A-38857
; Sequence 38857, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38857
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Salmonella typhimurium
US-10-282-122A-38857

Query Match 6.1%; Score 73; DB 7; Length 1134;
Best Local Similarity 45.6%; Pred. No. 0.00017;
Matches 296; Conservative 0; Mismatches 350; Indels 3; Gaps 1;

```
Qy        546 TTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGA 605
              ||||  | | |    || ||  |||| |||  || | |     | ||| |||     |||
Db        489 TTTAGTAGGTATTGATTATCATAAAACTATCGTTATCGAAGGTGCTGGAATTAACCCGGA 548

Qy        606 TCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGC 665
              | || || | || ||   || ||||  |    | ||||  |   |  | |||||  |
Db        549 TATTTATAAATTCTC---TATAGAGAAAAAACATGAGATACCGATAGTCCTTTTTGCAAG 605

Qy        666 AAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGAC 725
```

```
Db        606 CCGAATGTTATGGAGTAAAGGTCTTGGCGATCTCATTGAAGCAAAAAAGATCCTCAGACA 665

Qy        726 AACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTC 785
              ||   ||    || ||  ||    ||  |||     || |     | ||       |
Db        666 AAAAAATATTCATTTTGTTTTAAATGTTGCAGGTATATTGGCCGAGGACGATAAAGATGC 725

Qy        786 ATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCA 845
              | ||       |  ||| |          || |||       |||  |  |  || |
Db        726 TATCCCTTTGGAGCTCATTCATCACTGGCATAATGAAGGACTAATTAACTGGCTGGGACG 785

Qy        846 TGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATA 905
              |||||  | ||     | |||||    | |  |||  |||||  |     ||
Db        786 CTCTTCGAATGTATATGAGCTTATTCAGAAATCGAATATCGTTGCATTACCATCTATTTA 845

Qy        906 TCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAAC 965
              ||  ||||| |||||| || |  |  ||| ||  ||  || | ||  |  |   || |
Db        846 TCCGGAAGGAGTACCACGGCTTTTATTAGAGGCATCATCTGTAGGACGTGCATGTATTGC 905

Qy        966 AACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACC 1025
              |  ||     ||  |||       |||| | | ||      ||||| ||  |  |
Db        906 TTATGATACCGGCGGATGTGATAGCTTAATTATTCATAACTATAATGGGTTGATCGTGAA 965

Qy       1026 TCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAA 1085
              |  |  | |      | || | | | | | |||||  |  | |||  | | |
Db        966 AAGTAATTCAGCTCAGGAGTTAGCGGTAGAATTAGAGTATTTGTTAAAAAATCCACAGAT 1025

Qy       1086 AGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCTTTTGA 1145
              | || ||||||      ||||||  ||   || | |  ||   |||        |
Db       1026 TCGCTTAGAGATGGGGGCGAATGGAAGAAAACGTGTAAAGGAATTGTTTTCCAGCATATT 1085

Qy       1146 AAAAAATAATAGACTAGCATCAATAATAAAATCAAATAATGATTTTTGA 1194
              |   | |||||     |     |||     ||| | || || ||  | |
Db       1086 AGTCATTAATAAGACATTGCAAATTTATAAAGACACTATTGGTTGCTAA 1134
```

RESULT 5
US-10-282-122A-16045
; Sequence 16045, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

```
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16045
;   LENGTH: 1092
;   TYPE: DNA
;   ORGANISM: Clostridium botulinum
US-10-282-122A-16045

  Query Match             5.9%;  Score 70.2;  DB 7;  Length 1092;
  Best Local Similarity   45.5%;  Pred. No. 0.00059;
  Matches  329;  Conservative    0;  Mismatches  388;  Indels    6;  Gaps    2;

Qy        469 ATACAAATTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTATTAAAT 528
              ||| |||       |   | |||     ||| | |     |||  | | ||       | |
Db        361 ATAAAAAGAAAAATGGTTGAGTATAATTTATCCAAAGCAGATATTATAATGTCTACCAGT 420

Qy        529 CATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACAGTGTTA 588
               | |       | ||||   ||   |  |   || ||  ||      |  ||||    ||
Db        421 AAAGTAATGGCAGAAGAGACAAATCAATATACTACTAAAAATATAGAAATAACCCCTTTT 480

Qy        589 GGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCACCGAAAGAGAAAATT 648
              || |  | |  |       ||||    |   |  ||||| |   |   |||||  ||  |
Db        481 GGAGTAAATATAGATACATTTAAGCCTTGCGCCGATAAATATGAAAAAAAAGAAAATTTA 540

Qy        649 ACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAGCC 708
                  |   |     |  ||  |||        |  | || |||    | ||  ||  |||
Db        541 GTAATAGGTACAGTTAAAACATTAGAGCCTAAATATGGTATAGAATACTTAGTTAGAGCT 600

Qy        709 GCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTGAG 768
                  || |||  ||   ||   | |  | | || | ||||  |||||     ||
Db        601 TTTGCAAAGGTAAAGCAAAAACATAATAATATTAAGTTAGAAATTGCAGGAGTAGGGGAC 660

Qy        769 AGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATGAT 828
                ||  |   |||       | |   ||||| || |     ||| | |  |   | ||
Db        661 CAAAAGGACTTTTTACTTAATCTATGTAATGAACTTAA---TATAAAAGACCATGTTAAA 717

Qy        829 CTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTTTT 888
                || |       |    ||   ||||| ||   |||    ||  | | |  |  |||   |
Db        718 TTTTTAGGATTTATAAATCAAGAAAAGGTAATAGAGGCTTTTAATAGATTTGATGTAGCT 777

Qy        889 GTTTTACCT---ACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCT 945
              || || |||   |||| | || | ||| |  |       |          ||||||   |||
Db        778 GTGTTTCCTTCCACATTAGATAGTGAAAGTTTCGGGGTAGCAGCAGTAGAAGCACAGGCA 837

Qy        946 ATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATGGG 1005
               ||| |  || |||||     ||||||||   || |     ||    | |  |  |
Db        838 TGTGGGACTCCAGTAATTGTTTCTAATGTAGGAGGATTACCTGAAGCTACAGCTCCTAAT 897

Qy       1006 GTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAATAT 1065
                | | |  | ||  ||  |  |        |  ||   || |||||    ||  || |
Db        898 AATAGTAGTATATTAGTAAATAAAAAAAGTGTAGATGAGCTAGCAGAGGCCATAGAAAAA 957

Qy       1066 TTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAGAA 1125
              || ||| || || |  ||  |     |  | |||||      | ||||| || ||||  |||
Db        958 TTAATTAAGGATGACAATTTAAGGATAAACATGGGTAAGACAGGAAGAAAATTTGTGGAA 1017

Qy       1126 AAAAACTTTGATGCTTTTGAAAAAAATAATAGACTAGCATCAATAATAAAATCAAATAAT 1185
                | || ||| ||      || ||   |||||   |||    ||||   |||||  |  |
Db       1018 GATAATTTTAATATAGAAGATAATTTTAATAATGTAGATATAATATATAAATCTATAATA 1077

Qy       1186 GAT 1188
                |||
Db       1078 GAT 1080


RESULT 6
US-10-369-493-24115
; Sequence 24115, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 24115
;   LENGTH: 1080
;   TYPE: DNA
;   ORGANISM: Xenorhabdus nematophilus
US-10-369-493-24115

  Query Match             5.7%;  Score 67.8;  DB 6;  Length 1080;
  Best Local Similarity   49.6%;  Pred. No. 0.0018;
  Matches  174;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        824 ATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTG 883
              ||  | |||| | || |||  ||  | |  |||| |  || || ||| |   ||    |
Db        728 ATAGTGTTATCTTTCTTGGCAATCGGCAGGATGTGCCGGACTGTTTAAACTCAATGGATA 787

Qy        884 TTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGG 943
              ||||||  || ||| | |     |  ||||| || ||     |  ||      | || ||||
Db        788 TTTTTGCATTGCCTTCTTTTGGTAACGAAGGAGTGCCGCAAGGTATTATGCAGGCAATGG 847

Qy        944 CTATTGGTAGACCTGTAATAACAACTAATGTACCTGGGTGTAGGGATATAATAAATGATG 1003
              |   ||||  ||| || ||| | ||     ||     | |  |   ||      | | |||||
Db        848 CCTGTGGTTTACCGGTCATATCGACATCGGTTGGCGCGATTTCTGAGGCTGTGATTGATG 907

Qy       1004 GGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAAT 1063
              | |  | ||||| | | ||    ||       || |   | |  | ||  |  |||
Db        908 GTGAAACTGGCTATATTATCGAGCCTAAGTGTGTTGAACAATTAACTAATAAGTTAGAAT 967

Qy       1064 ATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCAG 1123
                | ||    |||  |||   |     |  | |||||   |||   |  |||    |||
Db        968 TACTAATAAGAAATGCAGAACTACGTTTACAGATGGGAGATGCCTCACTGAAACGGGCAA 1027

Qy       1124 AAAAAAACTTTGATGCTTTTGAAAAAAATAATAGACTAGCATCAATAATAA 1174
               |    | |||| |     | | |      ||| | | | ||| ||| | |
Db       1028 TAGCTGAATTTGGTATGGATAATATGCTGGATAAAATGGAATCTATATTTA 1078


RESULT 7
US-10-282-122A-17194
; Sequence 17194, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17194
; LENGTH: 1089
; TYPE: DNA
; ORGANISM: Clostridium difficile
US-10-282-122A-17194

Query Match 5.5%; Score 65.6; DB 7; Length 1089;
Best Local Similarity 45.1%; Pred. No. 0.0049;
Matches 285; Conservative 0; Mismatches 344; Indels 3; Gaps 1;

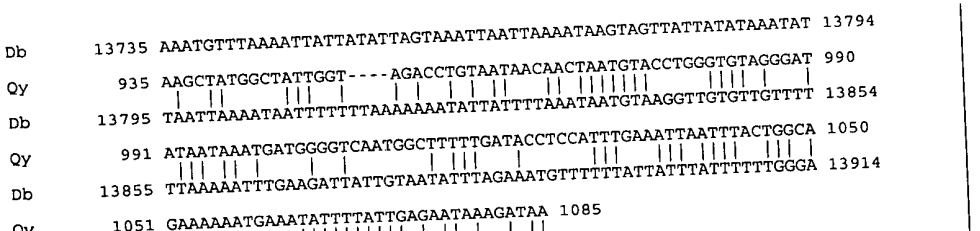
```
Qy       560 ATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTT 619
Db       446 ATAAATATACAGCTAAACATGTAGATGTAACCTTTTTTGGAGTAGATATGGATAGGTTTA 505
Qy       620 CATATAAAGAGCCACCGAAAGAGAAAATTACCTTTAT---TTTTATAGCAAGGTTATTAA 676
Db       506 AGCCTATGGAAGTAGAAAAAGAAGACGCTTTTGTAATAGGAATTATAAAGAGTTTAGAAA 565
Qy       677 GAGAGAAAGGGATATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAA 736
Db       566 AAAAATATGGAATAGAATATTTAATACAAGCATTTAAGATGCTTAAAGATGAATATAAAG 625
Qy       737 GTTCTGAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAA 796
Db       626 ATAAGAAAATAATTCTTAAAATAGGTGGCTCTGGAAGTCAGATGGACAACCTAATAAATT 685
Qy       797 ATGAAATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAAAATG 856
Db       686 TAGCTAAAGAATTGGGTATAGAAAATGATGTTCAGTTTTTAGGAAGAATATCTCCAGAAA 745
Qy       857 TTCAAGATTGGTTAGAGAAAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAAGGCG 916
Db       746 ATGTAAGCAAGACTTTTAATTCTTTTGATGTAACTGTATTTCCTTCTTTAAGAGAGGGAT 805
Qy       917 TACCAAGGGTGATCCAAGAAGCTATGGCTATTGGTAGACCTGTAATAACAACTAATGTAC 976
Db       806 TTGGAGTGGCAGCTATAGAATCAGAAGCATGTGAAGTACCTGTTATAGTTACAAATGTAG 865
Qy       977 CTGGGTGTAGGGATATAATAAATGATGGGGTCAATGGCTTTTTGATACCTCCATTTGAAA 1036
Db       866 GAGGACATCCAGAATCTGTATGGGAAAATGAAACTGGCTTAATTGTAGAACCAAAACAAC 925
Qy      1037 TTAATTTACTGGCAGAAAAAATGAAATATTTTATTGAGAATAAAGATAAAGTACTCGAAA 1096
Db       926 CAGAAGAAATAAAAAATGCTATAATAAAACTAATGGAAAATGATGAACTTAGACTTAATA 985
Qy      1097 TGGGGCTTGCTGGAAGGAAGTTTGCAGAAAAAAACTTTGATGCTTTTGAAAAAAATAATA 1156
Db       986 TGGGTAAGAAAGGGCGTCAATTTGTTAGAGAAAATTATGAGGTTAATTTAAACTTTAATG 1045
Qy      1157 GACTAGCATCAATAATAAAATCAAATAATGAT 1188
```

Db 1046 ATATAGAAAAAATTTATGATTCTATATTTGAT 1077

RESULT 8
US-10-311-455-597
; Sequence 597, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 597
; LENGTH: 16217
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 12269
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-597

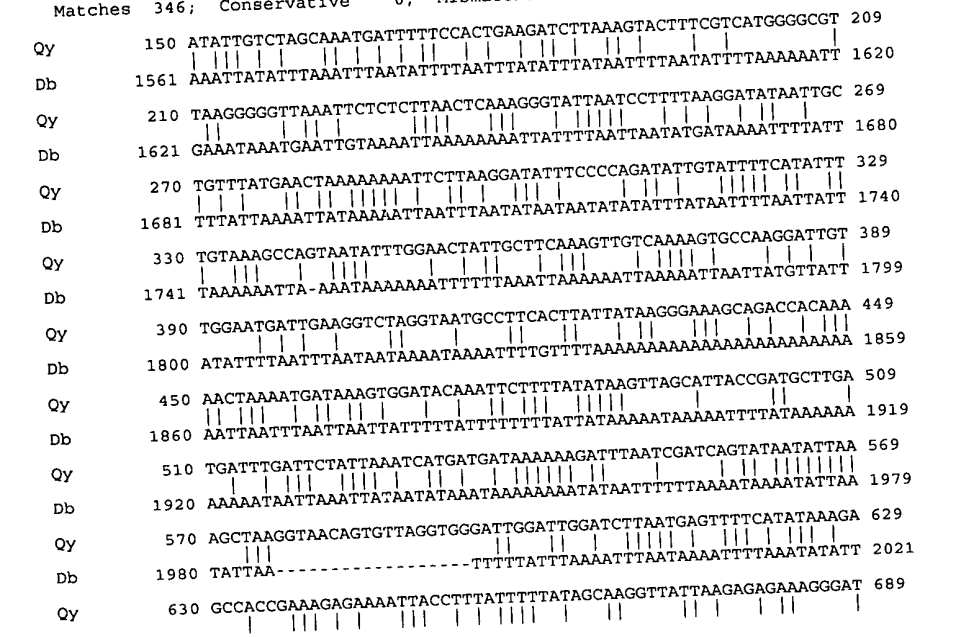
Query Match 5.0%; Score 59.8; DB 6; Length 16217;
Best Local Similarity 44.9%; Pred. No. 0.2;
Matches 312; Conservative 0; Mismatches 377; Indels 6; Gaps 2;

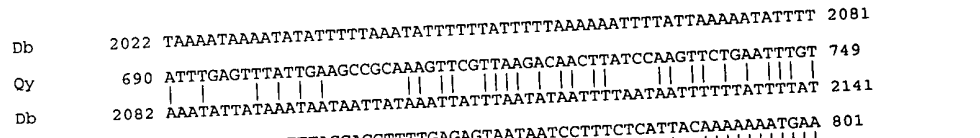
```
Qy       397 ATTGAAGGTCTAGGTAATGCCTTCACTTATTATAAGGGAAAGCAGACCACAAAAACTAAA 456
Db     13255 ATTTATTGTGTATATTATATTTAATAGGAATTTTAGTGATATAATTATATATATTTTT 13314
Qy       457 ATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATT-- 514
Db     13315 ATTAATAAATAGAGTATATGGATATTAAATTGTTTTATTTTATATATTTTTTAAATTTAA 13374
Qy       515 TGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAAAGCTA 574
Db     13375 TAATTAGATTTGTTTTTGATAATATTTGTAATTTATATTTATATTAAGGTATTAAGGTTA 13434
Qy       575 AGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCAC 634
Db     13435 TATTAATGTTTTAAAAATTATATAATTAAAATTTATAAAATAATTAGAAAATAAAAAGAA 13494
Qy       635 CGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTG 694
Db     13495 AGAAAAGTATAGTAATTTTTAAGATAAGGGAGGTTTTTTTTATGTTGGTATTTTTAATTT 13554
Qy       695 AGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTT 754
Db     13555 AGGTTATTAAAAGAGATAATTTAGTTTATTTAAATCGTTAGAAAATTATGTAAATATTTA 13614
Qy       755 TAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAA 814
Db     13615 TTGGTTTGATTTAGTTATTTTATAGTTTATATATATTTTAAAATAATATGTATATGATAA 13674
Qy       815 GAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGA 874
Db     13675 ATATATATAATTTTGTTAATGAAAATAGATTTATAAATAAAAATATATAAATTGTTTTTA 13734
Qy       875 AAAGTTCTGTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAG 934
```



RESULT 9
US-10-473-126-386
; Sequence 386, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 386
; LENGTH: 8056
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-386

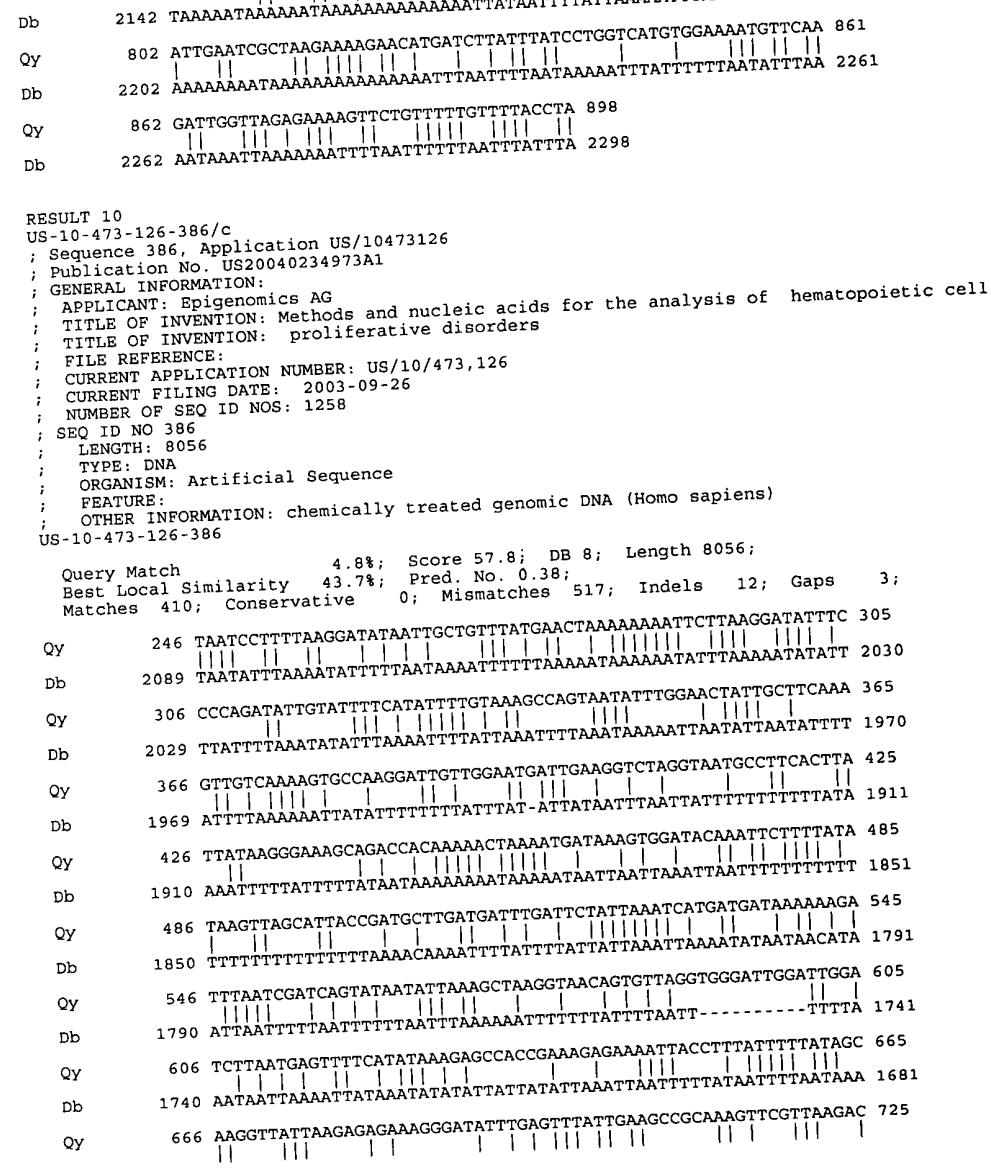
Query Match 4.9%; Score 58.6; DB 8; Length 8056;
Best Local Similarity 45.7%; Pred. No. 0.27;
Matches 346; Conservative 0; Mismatches 384; Indels 27; Gaps 3;

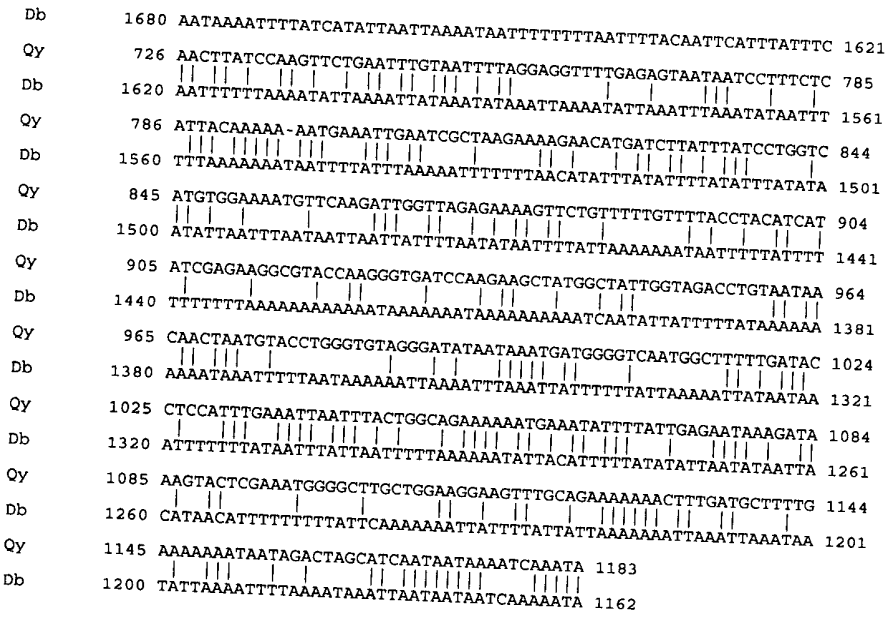




RESULT 10
US-10-473-126-386/c
; Sequence 386, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 386
; LENGTH: 8056
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-386

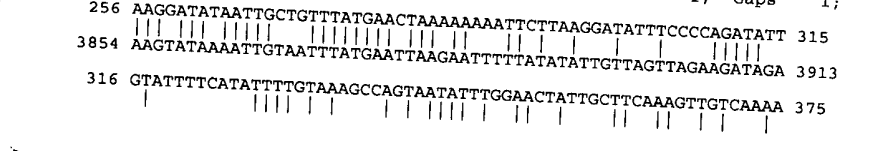
Query Match 4.8%; Score 57.8; DB 8; Length 8056;
Best Local Similarity 43.7%; Pred. No. 0.38;
Matches 410; Conservative 0; Mismatches 517; Indels 12; Gaps 3;





RESULT 11
US-10-311-455-531
; Sequence 531, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 531
; LENGTH: 5999
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-531

Query Match 4.8%; Score 57.6; DB 6; Length 5999;
Best Local Similarity 47.8%; Pred. No. 0.38;
Matches 197; Conservative 0; Mismatches 214; Indels 1; Gaps 1;

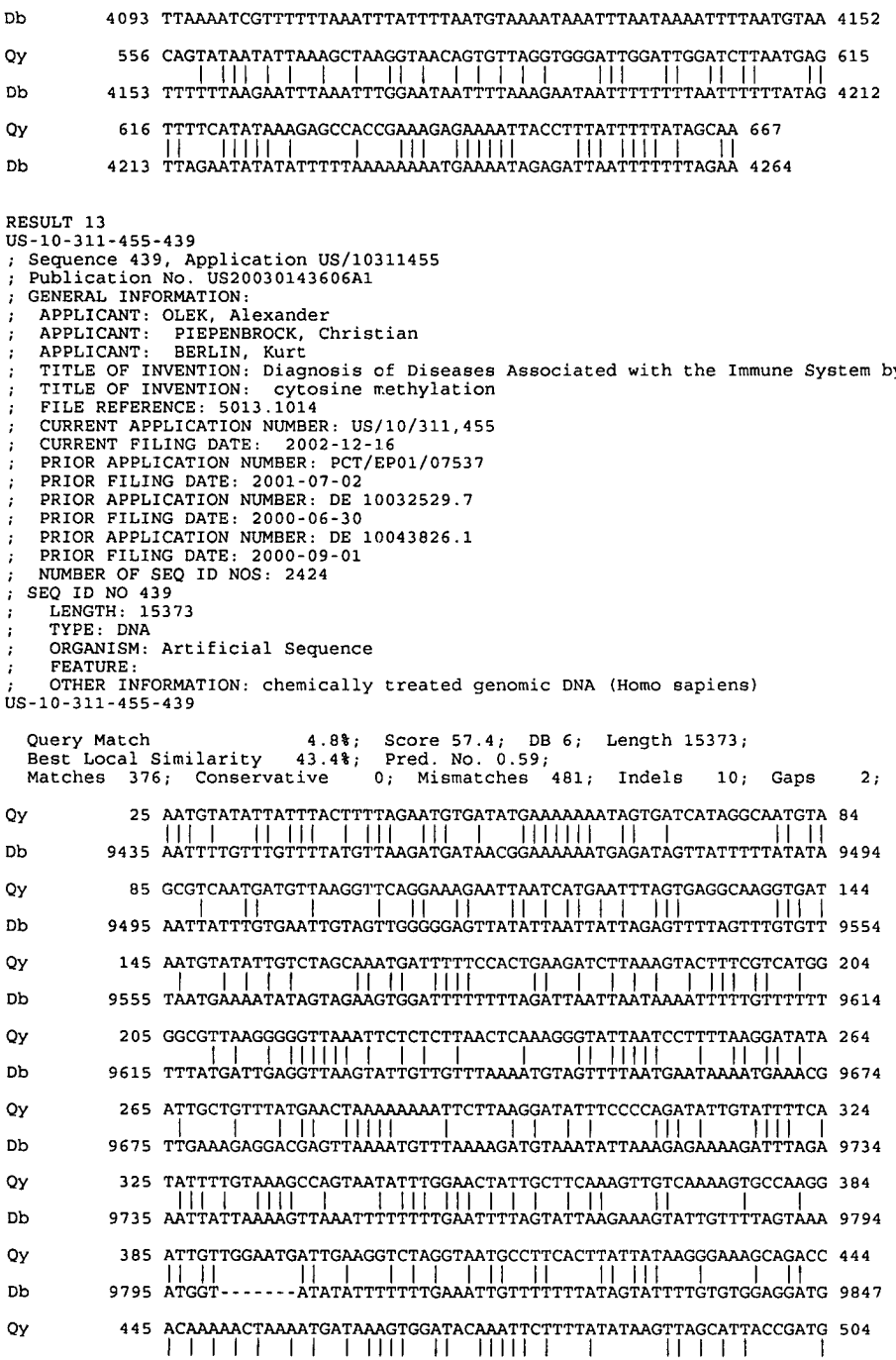




RESULT 12
US-10-257-166-41
; Sequence 41, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 41
; LENGTH: 5999
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-41

Query Match 4.8%; Score 57.6; DB 7; Length 5999;
Best Local Similarity 47.8%; Pred. No. 0.38;
Matches 197; Conservative 0; Mismatches 214; Indels 1; Gaps 1;





```
Db       4093 TTAAAATCGTTTTTTAAATTTATTTTAATGTAAAATAAATTTAATAAAATTTTAATGTAA 4152

Qy        556 CAGTATAATATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAG 615

Db       4153 TTTTTTAAGAATTTAAATTTGGAATAATTTTAAAGAATAATTTTTTTTAATTTTTTATAG 4212

Qy        616 TTTTCATATAAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAA 667

Db       4213 TTAGAATATATATTTTTAAAAAAAATGAAAATAGAGATTAATTTTTTTAGAA 4264


RESULT 13
US-10-311-455-439
; Sequence 439, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 439
;   LENGTH: 15373
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-439

  Query Match             4.8%;  Score 57.4;  DB 6;  Length 15373;
  Best Local Similarity   43.4%;  Pred. No. 0.59;
  Matches  376;  Conservative    0;  Mismatches  481;  Indels   10;  Gaps    2;

Qy         25 AATGTATATTATTTACTTTTAGAATGTGATATGAAAAAAATAGTGATCATAGGCAATGTA 84

Db       9435 AATTTTGTTTGTTTTATGTTAAGATGATAACGGAAAAAATGAGATAGTTATTTTTATATA 9494

Qy         85 GCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATCATGAATTTAGTGAGGCAAGGTGAT 144

Db       9495 AATTATTTGTGAATTGTAGTTGGGGGAGTTATATTAATTATTAGAGTTTTAGTTTGTGTT 9554

Qy        145 AATGTATATTGTCTAGCAAATGATTTTTCCACTGAAGATCTTAAAGTACTTTCGTCATGG 204

Db       9555 TAATGAAAATATAGTAGAAGTGGATTTTTTTTAGATTAATTAATAAAATTTTTGTTTTTT 9614

Qy        205 GGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTTAAGGATATA 264

Db       9615 TTTATGATTGAGGTTAAGTATTGTTGTTTAAAATGTAGTTTTAATGAATAAAATGAAACG 9674

Qy        265 ATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGATATTTCCCCAGATATTGTATTTTCA 324

Db       9675 TTGAAAGAGGACGAGTTAAAATGTTTAAAAGATGTAAATATTAAAGAGAAAAGATTTAGA 9734

Qy        325 TATTTTGTAAAGCCAGTAATATTTGGAACTATTGCTTCAAAGTTGTCAAAAGTGCCAAGG 384

Db       9735 AATTATTAAAAGTTAAATTTTTTTTGAATTTAGTATTAAGAAAGTATTGTTTTAGTAAA 9794

Qy        385 ATTGTTGGAATGATTGAAGGTCTAGGTAATGCCTTCACTTATTATAAGGGAAAGCAGACC 444

Db       9795 ATGGT-------ATATATTTTTTTGAAATTGTTTTTTTATAGTATTTTGTGTGGAGGATG 9847

Qy        445 ACAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATG 504

Db       9848 ATAGAGATTGTATTTTTTAAGTAAATTAAAATTATAGAAATATTTTTTGGAGTTGTATAG 9907

Qy        505 CTTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAAT 564

Db       9908 TATGTTTATAACGTTTAGTTTTATTTAGTGAAAATTTTTGATGAATTTAGATGTATATT 9967

Qy        565 ATTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATAT 624

Db       9968 AATATGAATTATTTAGTAATTATATTTTTTGAAAAATAGTGTGTTAATGAGATATTTTAA 10027

Qy        625 AAAGAGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAA 684

Db      10028 ATATATTTTTGTATTGGATGTAGAAGTTTTTTTTGTTTGGTATTATTGTTGTTAGAATTT 10087

Qy        685 GGGATATTTGAGTTTATTGAA---GCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCT 741

Db      10088 TATTTTTTTGTATATAGTAAAATGTATGTAGGGTTTTTGAGTTATATATTTATTTTGTTG 10147

Qy        742 GAATTTGTAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAA 801

Db      10148 GTGGTTTGGATGTTATTAGGGTAGTTTTAGAGATATAATTTTTTTTTTATAAGGTTAAGAG 10207

Qy        802 ATTGAATCGCTAAGAAAAGAACATGATCTTATTTATCCTGGTCATGTGGAAAATGTTCAA 861

Db      10208 AAAATAAATAGGAGAATTTAATTTTAGATGGTATTTTTTTGTTTTGAGGTTAGATAGAAA 10267

Qy        862 GATTGGTTAGAGAAAAGTTCTGTTTTT 888

Db      10268 AATATTTTGTAAAGAATTGTTGTTTTT 10294


RESULT 14
US-10-278-698-255
; Sequence 255, Application US/10278698
; Publication No. US20050037344A1
; GENERAL INFORMATION:
;  APPLICANT: PathoArray GmbH
;  APPLICANT:  Stuhlmuller, Bruno
;  APPLICANT:  Haupl, Thomas
;  TITLE OF INVENTION: Nucleic Acid Array
;  FILE REFERENCE: O30027US
;  CURRENT APPLICATION NUMBER: US/10/278,698
;  CURRENT FILING DATE:  2002-10-23
;  NUMBER OF SEQ ID NOS: 1050
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 255
;   LENGTH: 115218
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-278-698-255

  Query Match             4.8%;  Score 56.8;  DB 8;  Length 115218;
  Best Local Similarity   46.6%;  Pred. No. 1.7;
  Matches  320;  Conservative    0;  Mismatches  357;  Indels    9;  Gaps    4;

Qy         17 TACATATCAATGTATATTATTTACTTTTAGAATGTGATATGAAAAAAATAGTGATCATAG 76

Db      16557 TAATTATCAGTAATAATTATATAATTATTATATATTATATATAAGTAATTATCAGTAAAG 16616

Qy         77 GCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAATTAATCATGAATTTAGTGAGGC 136

Db      16617 TGGAAGTAATAATAAAGGAAGTAATATAATATAATAAGTATATAAATATATAATATAATA 16676

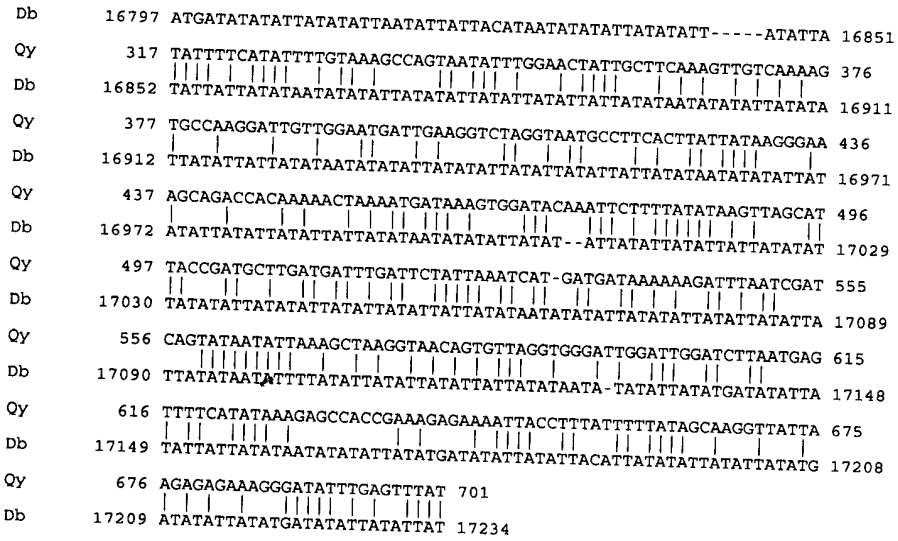
Qy        137 AAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCACTGAAGATCTTAAAGTACTTT 196

Db      16677 TAATATATTATGAATATATTATATTATATTATTATTATATAATATATATTATATATTGTA 16736

Qy        197 CGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACTCAAAGGGTATTAATCCTTTTA 256

Db      16737 TTATATAATTATTATATATATTATATTATTATATAATTATATTATATATTATATTATTAT 16796

Qy        257 AGGATATAATTGCTGTTTATGAACTAAAAAAAATTCTTAAGGATATTTCCCCAGATATTG 316
```



RESULT 15
US-10-278-698-769
; Sequence 769, Application US/10278698
; Publication No. US20050037344A1
; GENERAL INFORMATION:
; APPLICANT: PathoArray GmbH
; APPLICANT: Stuhlmuller, Bruno
; APPLICANT: Haupl, Thomas
; TITLE OF INVENTION: Nucleic Acid Array
; FILE REFERENCE: O30027US
; CURRENT APPLICATION NUMBER: US/10/278,698
; CURRENT FILING DATE: 2002-10-23
; NUMBER OF SEQ ID NOS: 1050
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 769
; LENGTH: 115218
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-278-698-769

Query Match 4.8%; Score 56.8; DB 8; Length 115218;
Best Local Similarity 46.6%; Pred. No. 1.7;
Matches 320; Conservative 0; Mismatches 357; Indels 9; Gaps 4;





Search completed: May 11, 2006, 07:09:12
Job time : 795.075 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:23:58 ; Search time 617.001 Seconds
(without alignments)
7892.621 Million cell updates/sec

Title: US-09-423-093B-1_COPY_739_1932
Perfect score: 1194
Sequence: 1 atgttactatgctgcataca.....aatcaaataatgatttttga 1194

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	217.8	18.2	1227	10	US-10-467-657-4219	Sequence 4219, Ap
2	217.8	18.2	1227	10	US-10-467-657-7087	Sequence 7087, Ap
3	152.8	12.8	1158	18	US-11-079-463-917	Sequence 917, App
4	144.8	12.1	1131	18	US-11-079-463-2639	Sequence 2639, Ap
5	49.8	4.2	1191	17	US-11-098-686-9334	Sequence 9334, Ap
c 6	49.8	4.2	1457619	17	US-11-098-686-8739	Sequence 8739, Ap
c 7	49	4.1	637	12	US-10-301-480-404165	Sequence 404165,
c 8	49	4.1	637	12	US-10-301-480-1017574	Sequence 1017574,
c 9	49	4.1	638	7	US-09-925-065A-330315	Sequence 330315,
c 10	46.4	3.9	626	11	US-10-301-480-79342	Sequence 79342, A
c 11	46.4	3.9	626	12	US-10-301-480-692751	Sequence 692751,
c 12	45.2	3.8	536	7	US-09-925-065A-312595	Sequence 312595,
c 13	45.2	3.8	536	7	US-09-925-065A-312596	Sequence 312596,
c 14	45.2	3.8	558	12	US-10-301-480-387826	Sequence 387826,
c 15	45.2	3.8	558	12	US-10-301-480-387827	Sequence 387827,
c 16	45.2	3.8	558	12	US-10-301-480-1001235	Sequence 1001235,
c 17	45.2	3.8	558	12	US-10-301-480-1001236	Sequence 1001236,
c 18	45.2	3.8	626	11	US-10-301-480-79340	Sequence 79340, A
c 19	45.2	3.8	626	11	US-10-301-480-79341	Sequence 79341, A
c 20	45.2	3.8	626	12	US-10-301-480-692749	Sequence 692749,
c 21	45.2	3.8	626	12	US-10-301-480-692750	Sequence 692750,
22	45.2	3.8	1095	11	US-10-932-182A-150	Sequence 150, App
23	45.2	3.8	1095	11	US-10-932-182A-150	Sequence 150, App
24	45.2	3.8	1107	18	US-11-079-463-1627	Sequence 1627, Ap
25	45.2	3.8	186442	17	US-11-121-086-104	Sequence 104, App
c 26	44.6	3.7	171486	17	US-11-121-086-105	Sequence 105, App
c 27	44.4	3.7	5026	11	US-10-932-182A-166180	Sequence 166180,
c 28	44.4	3.7	5026	11	US-10-932-182A-166180	Sequence 166180,
c 29	44	3.7	558	12	US-10-301-480-418366	Sequence 418366,
c 30	44	3.7	558	12	US-10-301-480-418368	Sequence 418368,
c 31	44	3.7	558	12	US-10-301-480-1031775	Sequence 1031775,
c 32	44	3.7	558	12	US-10-301-480-1031777	Sequence 1031777,
33	43.8	3.7	994	12	US-10-301-480-611274	Sequence 611274,
34	43.8	3.7	994	12	US-10-301-480-1224683	Sequence 1224683,
35	43.8	3.7	1269	10	US-10-336-263A-3	Sequence 3, Appli
c 36	43.4	3.6	1808	7	US-09-925-065A-82775	Sequence 82775, A
c 37	43.4	3.6	1808	11	US-10-301-480-184015	Sequence 184015,
c 38	43.4	3.6	1808	12	US-10-301-480-797424	Sequence 797424,
39	43.4	3.6	171486	17	US-11-121-086-105	Sequence 105, App
40	43.2	3.6	6499	17	US-11-011-332A-90	Sequence 90, Appl
c 41	43.2	3.6	1082144	17	US-11-117-187-211	Sequence 211, App
42	43	3.6	1113	18	US-11-079-463-145	Sequence 145, App
43	43	3.6	19233	10	US-10-240-708-46	Sequence 46, Appl
c 44	42.8	3.6	1775	10	US-10-750-185-48024	Sequence 48024, A
c 45	42.8	3.6	1775	10	US-10-750-623-48024	Sequence 48024, A

ALIGNMENTS

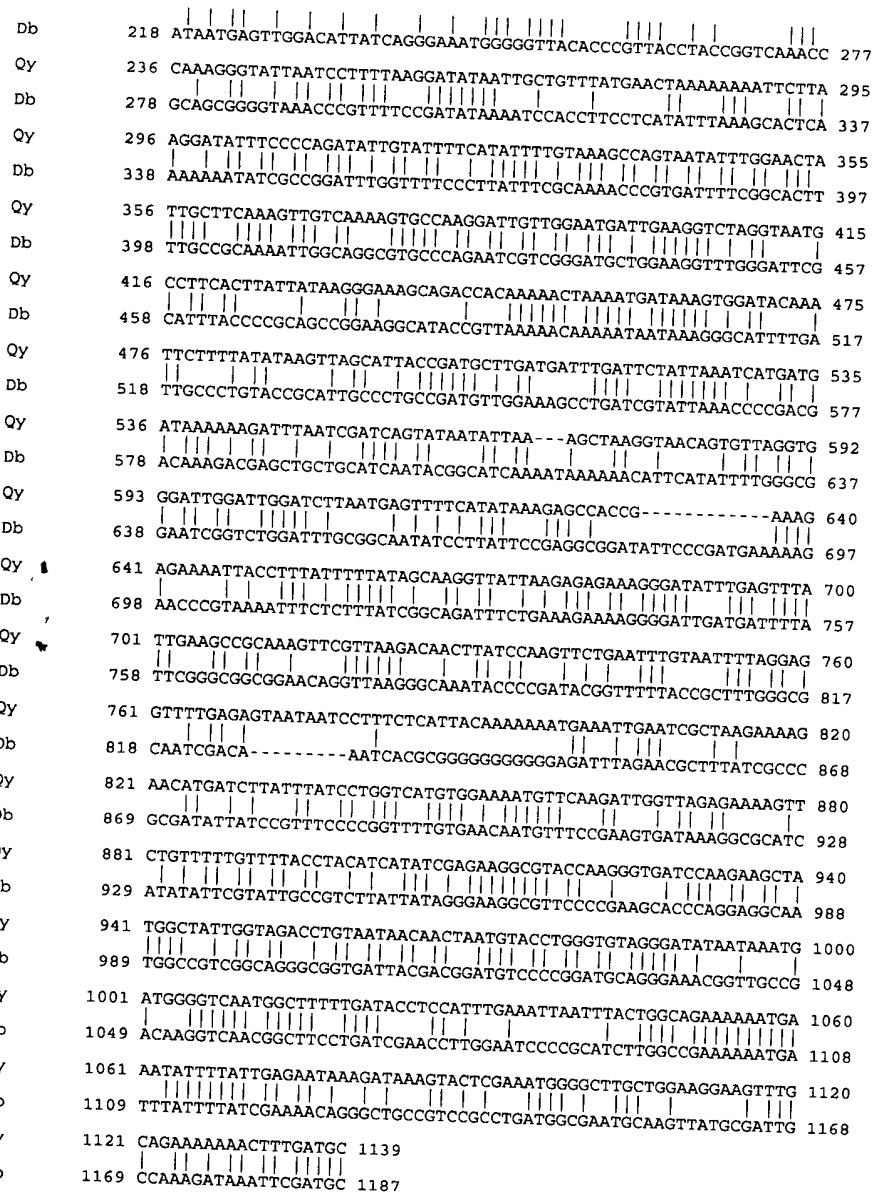
RESULT 1
US-10-467-657-4219
; Sequence 4219, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 4219
; LENGTH: 1227
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-4219

Query Match 18.2%; Score 217.8; DB 10; Length 1227;
Best Local Similarity 52.1%; Pred. No. 6.2e-30;
Matches 573; Conservative 0; Mismatches 502; Indels 24; Gaps 3;

```
Qy      56 TGAAAAAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGGTTCAGGAAAGAAT 115
           | |  ||||| ||  | ||     |  || || || |  || |  |  | ||| |
Db      98 TTATGAAAATCGTTTTTATCACAACAGTCGCATCCAGCATTTACGGTTTCCGCGCCCCCG 157

Qy     116 TAATCATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCAAATGATTTTTCCA 175
           | || |  || ||| |  |  ||     || | || |||     | | |   || ||||||
Db     158 TCATTAAAAAATTAATCGGCAAAAACCATCAGGTGTATGCCTTTGTATCGGAGTTTTCCG 217

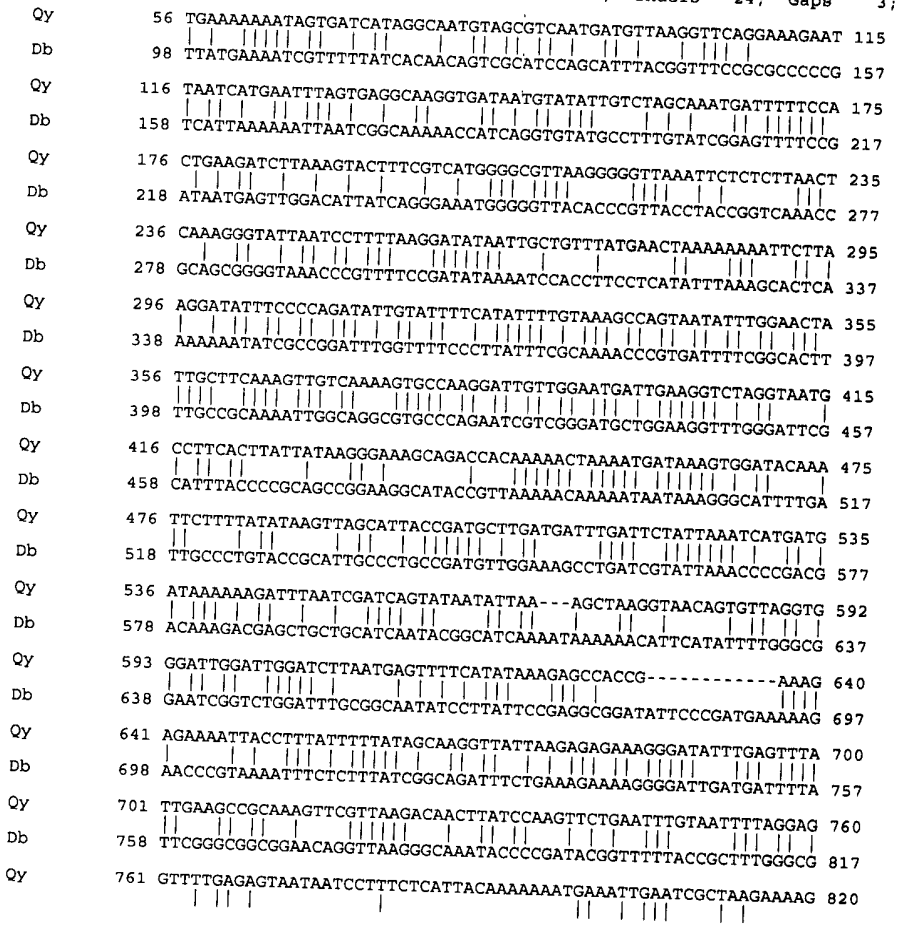
Qy     176 CTGAAGATCTTAAAGTACTTTCGTCATGGGGCGTTAAGGGGGTTAAATTCTCTCTTAACT 235
```

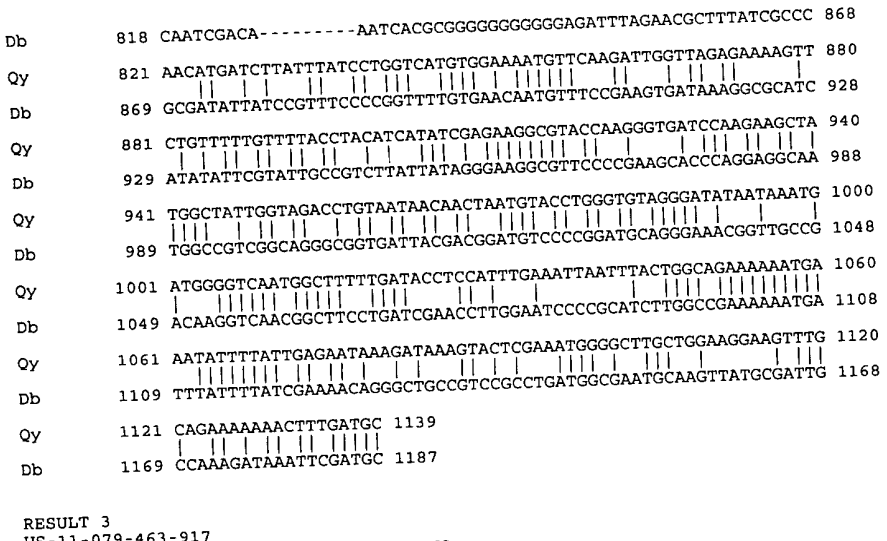


RESULT 2
US-10-467-657-7087
; Sequence 7087, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:

; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7087
; LENGTH: 1227
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7087

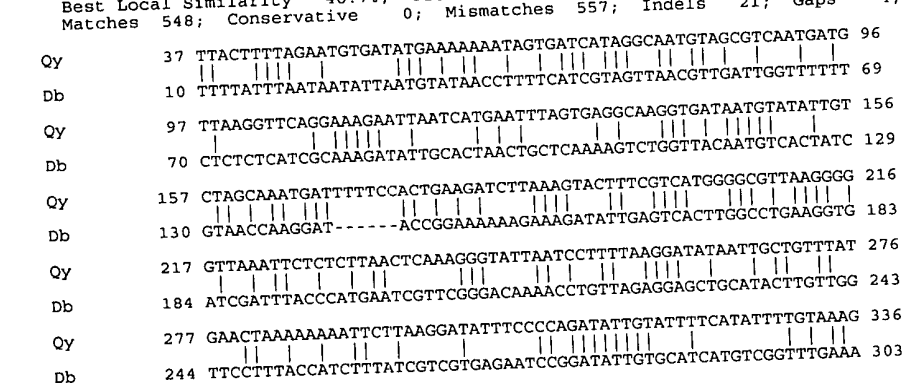
Query Match 18.2%; Score 217.8; DB 10; Length 1227;
Best Local Similarity 52.1%; Pred. No. 6.2e-30;
Matches 573; Conservative 0; Mismatches 502; Indels 24; Gaps 3;

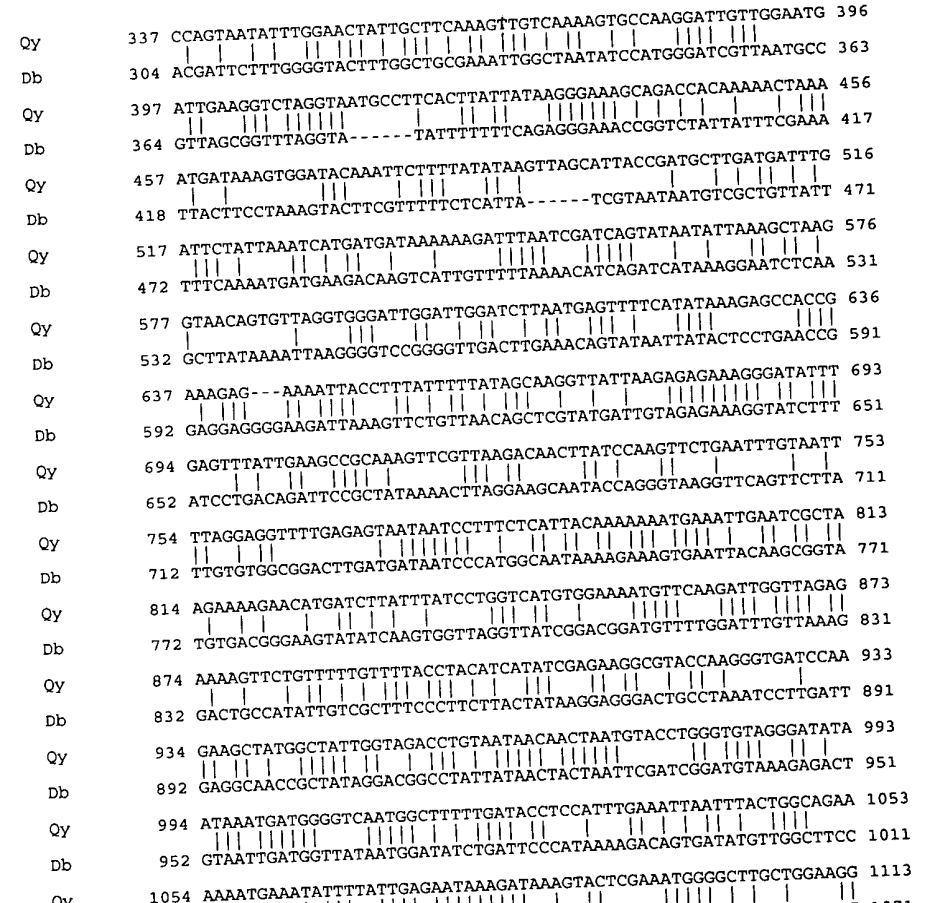




RESULT 3
US-11-079-463-917
; Sequence 917, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 917
; LENGTH: 1158
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-917

Query Match 12.8%; Score 152.8; DB 18; Length 1158;
Best Local Similarity 48.7%; Pred. No. 3.1e-18;
Matches 548; Conservative 0; Mismatches 557; Indels 21; Gaps 4;

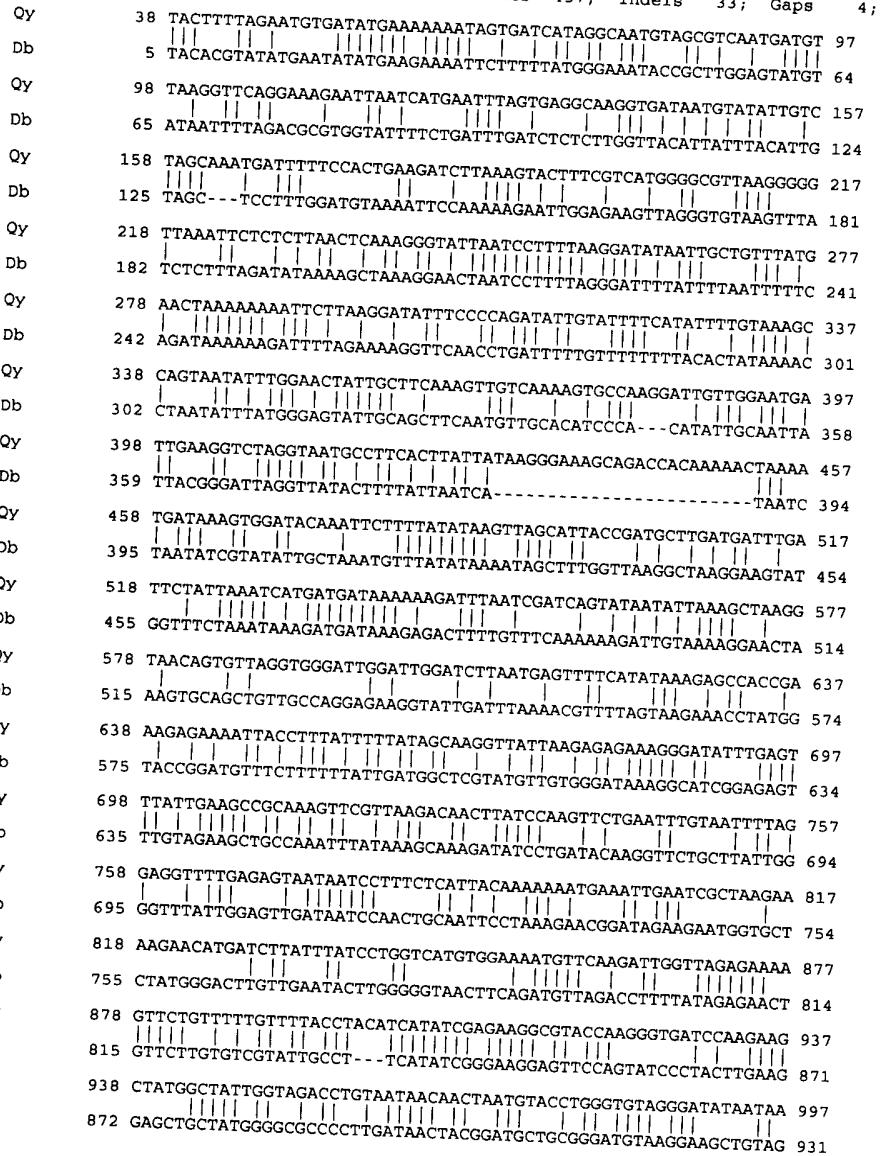




RESULT 4
US-11-079-463-2639
; Sequence 2639, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAC
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444

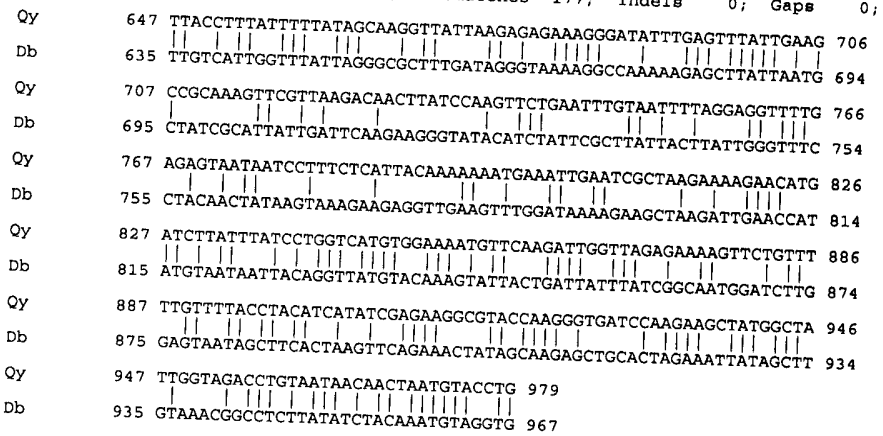
; SEQ ID NO 2639
; LENGTH: 1131
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-2639

Query Match 12.1%; Score 144.8; DB 18; Length 1131;
Best Local Similarity 50.1%; Pred. No. 8.4e-17;
Matches 492; Conservative 0; Mismatches 457; Indels 33; Gaps 4;



RESULT 5
US-11-098-686-9334
; Sequence 9334, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9334
; LENGTH: 1191
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-9334

Query Match 4.2%; Score 49.8; DB 17; Length 1191;
Best Local Similarity 46.8%; Pred. No. 11;
Matches 156; Conservative 0; Mismatches 177; Indels 0; Gaps 0;



RESULT 6
US-11-098-686-8739/c
; Sequence 8739, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04

```
;  NUMBER OF SEQ ID NOS: 11433
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8739
;    LENGTH: 1457619
;    TYPE: DNA
;    ORGANISM: Lawsonia intracellularis
US-11-098-686-8739

  Query Match             4.2%;  Score 49.8;  DB 17;  Length 1457619;
  Best Local Similarity   46.8%;  Pred. No. 27;
  Matches  156;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        647 TTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATTGAAG 706
              ||  | ||  |||  |||   | ||    || |  |||||   |   ||| ||||| | |
Db     529704 TTGTCATTGGTTTATTAGGGCGCTTTGATAGGGTAAAAGGCCAAAAAGAGCTTATTAATG 529645

Qy        707 CCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTGTAATTTTAGGAGGTTTTG 766
              |        ||  | |    |         |  |||       || |  |    || |||
Db     529644 CTATCGCATTATTGATTCAAGAAGGGTATACATCTATTCGCTTATTACTTATTGGGTTTC 529585

Qy        767 AGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAATCGCTAAGAAAAGAACATG 826
                | | ||    |     |        ||  |   ||  ||      | |  ||||
Db     529584 CTACAACTATAAGTAAAGAAGAGGTTGAAGTTTGGATAAAAGAAGCTAAGATTGAACCAT 529525

Qy        827 ATCTTATTTATCCTGGTCATGTGGAAAATGTTCAAGATTGGTTAGAGAAAAGTTCTGTTT 886
              || | ||   | | ||| |'||  ||| | ||   ||||  |||  |  ||    | ||
Db     529524 ATGTAATAATTACAGGTTATGTACAAAGTATTACTGATTATTTATCGGCAATGGATCTTG 529465

Qy        887 TTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATGGCTA 946
                ||  || || ||   |  |  ||||    || |||| |     | ||||  ||| |||
Db     529464 GAGTAATAGCTTCACTAAGTTCAGAAACTATAGCAAGAGCTGCACTAGAAATTATAGCTT 529405

Qy        947 TTGGTAGACCTGTAATAACAACTAATGTACCTG 979
                |    | ||| | ||| | || ||||||  ||
Db     529404 GTAAACGGCCTCTTATATCTACAAATGTAGGTG 529372


RESULT 7
US-10-301-480-404165/c
; Sequence 404165, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 404165
;    LENGTH: 637
;    TYPE: DNA
;    ORGANISM: Homo sapien
US-10-301-480-404165

  Query Match             4.1%;  Score 49;  DB 12;  Length 637;
  Best Local Similarity   49.8%;  Pred. No. 14;
  Matches  124;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy        455 AAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATT 514
              || |  ||||    ||  ||| || |    |||| | |  | |     || || ||  ||
Db        343 AATTATTAAAAATAATTAAAAATCATAATCATAAATAATAAATTATTTTGATTTATTTTT 284

Qy        515 TGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAAAGCTA 574
                  || | ||  || || |   |||| ||  | |  | ||||||  ||| ||
Db        283 AATAAATACTGAAAAATAATAAATTAAAAAATAAATTTTAACAGTATGATACTATTATAG 224

Qy        575 AGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCAC 634
```

```
              | |||| |   |||  | | |  | ||  ||   |||  || ||    | |   |   |
Db        223 AAGTAAAACAATTAATTAGAAAAGAATCAGAAGGTAACAAGATTCATAAAAGTAAATGAA 164

Qy        635 CGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTG 694
              |  ||   |||| |  | |||||||  | |  |   || ||   |   |      | | |
Db        163 CACAATTCAAAACTGTCGTTATTTTAAAATTATAATTTTTCCTAATTATTTTCCTTTGTT 104

Qy        695 AGTTTATTG 703
                |||||||
Db        103 CATTTATTG 95


RESULT 8
US-10-301-480-1017574/c
; Sequence 1017574, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1017574
;    LENGTH: 637
;    TYPE: DNA
;    ORGANISM: Homo sapien
US-10-301-480-1017574

  Query Match             4.1%;  Score 49;  DB 12;  Length 637;
  Best Local Similarity   49.8%;  Pred. No. 14;
  Matches  124;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy        455 AAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATT 514
              || |  ||||    ||  ||| || |    |||| | |  | |     || || ||  ||
Db        343 AATTATTAAAAATAATTAAAAATCATAATCATAAATAATAAATTATTTTGATTTATTTTT 284

Qy        515 TGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAAAGCTA 574
                  || | ||  || || |   |||| ||  | |  | ||||||  ||| ||
Db        283 AATAAATACTGAAAAATAATAAATTAAAAAATAAATTTTAACAGTATGATACTATTATAG 224

Qy        575 AGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCAC 634
              | |||| |   |||  | | |  | ||  ||   |||  || ||    | |   |   |
Db        223 AAGTAAAACAATTAATTAGAAAAGAATCAGAAGGTAACAAGATTCATAAAAGTAAATGAA 164

Qy        635 CGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTG 694
              |  ||   |||| |  | |||||||  | |  |   || ||   |   |      | | |
Db        163 CACAATTCAAAACTGTCGTTATTTTAAAATTATAATTTTTCCTAATTATTTTCCTTTGTT 104

Qy        695 AGTTTATTG 703
                |||||||
Db        103 CATTTATTG 95


RESULT 9
US-09-925-065A-330315/c
; Sequence 330315, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
```

```
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 330315
;   LENGTH: 638
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-330315

  Query Match             4.1%;  Score 49;  DB 7;  Length 638;
  Best Local Similarity   49.8%;  Pred. No. 14;
  Matches  124;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy        455 AAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGCTTGATGATT 514
Db        344 AATTATTAAAAATAATTAAAAATCATAATCATAAATAATAAATTATTTTGATTTATTTTT 285

Qy        515 TGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAAAGCTA 574
Db        284 AATAAATACTGAAAAATAATAAATTAAAAAATAAATTTTAACAGTATGATACTATTATAG 225

Qy        575 AGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAAGAGCCAC 634
Db        224 AAGTAAAACAATTAATTAGAAAAGAATCAGAAGGTAACAAGATTCATAAAAGTAAATGAA 165

Qy        635 CGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTG 694
Db        164 CACAATTCAAAACTGTCGTTATTTTAAAATTATAATTTTTCCTAATTATTTTCCTTTGTT 105

Qy        695 AGTTTATTG 703
Db        104 CATTTATTG 96


RESULT 10
US-10-301-480-79342/c
; Sequence 79342, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 79342
;   LENGTH: 626
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-79342

  Query Match             3.9%;  Score 46.4;  DB 11;  Length 626;
  Best Local Similarity   46.5%;  Pred. No. 40;
  Matches  146;  Conservative    1;  Mismatches  167;  Indels    0;  Gaps    0;

Qy        509 ATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTA 568
Db        626 ATAATTGGAAAATCCACACATATGTGGAAATTGAACAACACATTTTTTAATTTAACCGAT 567

Qy        569 AAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAAG 628
Db        566 GTGCTAAAGAAGAAATCTCAAGGGAAATTAGAAAATAYTTTAAAACAAGTGAAAATGAAA 507

Qy        629 AGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGA 688
Db        506 ATACAATAAACCAACACTTTTGAAATGCAGCTAAAAATGTGTTATAAGGAGACATAGCTG 447

Qy        689 TATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTG 748
Db        446 TAACTTCATGTCTTCAAAAAGAAGAACTATATTAAATCACTACTTTAACTTCATACCTTA 387

Qy        749 TAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAAT 808
Db        386 AAATATTAGAAAAATAAAAGCAAACTAAACCCAAATCTAGTAGAAGGAAGGAAATAATAA 327

Qy        809 CGCTAAGAAAAGAA 822
Db        326 TGATTACAATAGCA 313


RESULT 11
US-10-301-480-692751/c
; Sequence 692751, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 692751
;   LENGTH: 626
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-692751

  Query Match             3.9%;  Score 46.4;  DB 12;  Length 626;
  Best Local Similarity   46.5%;  Pred. No. 40;
  Matches  146;  Conservative    1;  Mismatches  167;  Indels    0;  Gaps    0;

Qy        509 ATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTA 568
Db        626 ATAATTGGAAAATCCACACATATGTGGAAATTGAACAACACATTTTTTAATTTAACCGAT 567

Qy        569 AAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATAAAG 628
Db        566 GTGCTAAAGAAGAAATCTCAAGGGAAATTAGAAAATAYTTTAAAACAAGTGAAAATGAAA 507

Qy        629 AGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATTAAGAGAGAAAGGGA 688
Db        506 ATACAATAAACCAACACTTTTGAAATGCAGCTAAAAATGTGTTATAAGGAGACATAGCTG 447

Qy        689 TATTTGAGTTTATTGAAGCCGCAAAGTTCGTTAAGACAACTTATCCAAGTTCTGAATTTG 748
Db        446 TAACTTCATGTCTTCAAAAAGAAGAACTATATTAAATCACTACTTTAACTTCATACCTTA 387

Qy        749 TAATTTTAGGAGGTTTTGAGAGTAATAATCCTTTCTCATTACAAAAAAATGAAATTGAAT 808
Db        386 AAATATTAGAAAAATAAAAGCAAACTAAACCCAAATCTAGTAGAAGGAAGGAAATAATAA 327

Qy        809 CGCTAAGAAAAGAA 822
Db        326 TGATTACAATAGCA 313


RESULT 12
```

```
US-09-925-065A-312595/c
; Sequence 312595, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 312595
;   LENGTH: 536
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-312595

  Query Match             3.8%;  Score 45.2;  DB 7;  Length 536;
  Best Local Similarity   52.6%;  Pred. No. 64;
  Matches  121;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        446 CAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGC 505
              |||  | || ||  || ||   |||  || |  | |   |||| ||      |  | ||||
Db        479 CAATTAATAGAAGAATTAAAAAGATTTAATTGTTATGCCATAATTTGATGGTCACCATGC 420

Qy        506 TTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATA 565
              |  ||  ||||  |  ||||  ||||  |      |||  || ||  ||||  | | ||
Db        419 TAAATTCTTTGTATTAATTATCTCATTTTTGAGTTAAATCATTTATTTATCAACAAATTA 360

Qy        566 TTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATA 625
              ||| || | |  ||  | | |  ||| ||     |      | ||   ||  |||   |||
Db        359 TTATAGATTACCTACTACTCTGTGGTAGGCACTGTGGTACTGTTCTGGACATTTGGAATA 300

Qy        626 AAG-AGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATT 674
                |  ||  | | |||||||  |   |  | ||  || || | ||   ||||
Db        299 TATCAGTGAACAAAAGAGACACAAATTTCTACCTTAATGGAAATTATATT 250


RESULT 13
US-09-925-065A-312596/c
; Sequence 312596, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 312596
;   LENGTH: 536
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-312596

  Query Match             3.8%;  Score 45.2;  DB 7;  Length 536;
  Best Local Similarity   52.6%;  Pred. No. 64;
  Matches  121;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        446 CAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGC 505
              |||  | || ||  || ||   |||  || |  | |   |||| ||      |  | ||||
Db        479 CAATTAATAGAAGAATTAAAAAGATTTAATTGTTATGCCATAATTTGATGGTCACCATGC 420

Qy        506 TTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATA 565
              |  ||  ||||  |  ||||  ||||  |      |||  || ||  ||||  | | ||
Db        419 TAAATTCTTTGTATTAATTATCTCATTTTTGAGTTAAATCATTTATTTATCAACAAATTA 360

Qy        566 TTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATA 625
              ||| || | |  ||  | | |  ||| ||     |      | ||   ||  |||   |||
Db        359 TTATAGATTACCTACTACTCTGTGGTAGGCACTGTGGTACTGTTCTGGACATTTGGAATA 300

Qy        626 AAG-AGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATT 674
                |  ||  | | |||||||  |   |  | ||  || || | ||   ||||
Db        299 TATCAGTGAACAAAAGAGACACAAATTTCTACCTTAATGGAAATTATATT 250


RESULT 14
US-10-301-480-387826/c
; Sequence 387826, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 387826
;   LENGTH: 558
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-387826

  Query Match             3.8%;  Score 45.2;  DB 12;  Length 558;
  Best Local Similarity   52.6%;  Pred. No. 65;
  Matches  121;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        446 CAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGC 505
              |||  | || ||  || ||   |||  || |  | |   |||| ||      |  | ||||
Db        501 CAATTAATAGAAGAATTAAAAAGATTTAATTGTTATGCCATAATTTGATGGTCACCATGC 442

Qy        506 TTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATA 565
              |  ||  ||||  |  ||||  ||||  |      |||  || ||  ||||  | | ||
Db        441 TAAATTCTTTGTATTAATTATCTCATTTTTGAGTTAAATCATTTATTTATCAACAAATTA 382

Qy        566 TTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATA 625
              ||| || | |  ||  | | |  ||| ||     |      | ||   ||  |||   |||
Db        381 TTATAGATTACCTACTACTCTGTGGTAGGCACTGTGGTACTGTTCTGGACATTTGGAATA 322

Qy        626 AAG-AGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATT 674
                |  ||  | | |||||||  |   |  | ||  || || | ||   ||||
Db        321 TATCAGTGAACAAAAGAGACACAAATTTCTACCTTAATGGAAATTATATT 272


RESULT 15
US-10-301-480-387827/c
; Sequence 387827, Application US/10301480
```

; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 387827
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-387827

Query Match 3.8%; Score 45.2; DB 12; Length 558;
Best Local Similarity 52.6%; Pred. No. 65;
Matches 121; Conservative 0; Mismatches 108; Indels 1; Gaps 1;

```
Qy        446 CAAAAACTAAAATGATAAAGTGGATACAAATTCTTTTATATAAGTTAGCATTACCGATGC 505
              |||   | || ||  || ||   |||  || |  | |   |||| ||     |  | ||||
Db        501 CAATTAATAGAARAATTAAAAAGATTTAATTGTTATGCCATAATTTGATGGTCACCATGC 442

Qy        506 TTGATGATTTGATTCTATTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATA 565
              |  ||  |||| |  |||| |||| |      |||   || ||  |||| | | ||
Db        441 TAAATTCTTTGTATTAATTATCTCATTTTTGAGTTAAATCATTTATTTATCAACAAATTA 382

Qy        566 TTAAAGCTAAGGTAACAGTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTCATATA 625
              ||| || | |  ||  | | |  ||| ||    |       | ||   ||  |||   |||
Db        381 TTATAGATTACCTACTACTCTGTGGTAGGCACTGTGGTACTGTTCTGGACATTTGGAATA 322

Qy        626 AAG-AGCCACCGAAAGAGAAAATTACCTTTATTTTTATAGCAAGGTTATT 674
              |  ||  | | |||||||  |   |  | ||  || || | ||   ||||
Db        321 TATCAGTGAACAAAAGAGACACAAATTTCTACCTTAATGGAAATTATATT 272
```

Search completed: May 11, 2006, 08:30:35
Job time : 621.001 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:45:22 ; Search time 4595.32 Seconds
(without alignments)
15660.244 Million cell updates/sec

Title: US-09-423-093B-1_COPY_8646_9911
Perfect score: 1266
Sequence: 1 atggtattaacagtgaaaaa.....gagcgaaaaatgtctattga 1266

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1266	100.0	14516	1	AF078736	AF078736 Escherich
	2	1266	100.0	14516	6	BD085554	BD085554 Nucleic a
	3	1266	100.0	14520	6	AX029261	AX029261 Sequence
	4	906.6	71.6	13650	1	AF285969	AF285969 Salmonell
	5	101.2	8.0	1328	15	AJ840631	AJ840631 Arabidops
	6	98.4	7.8	223697	14	AC151103	AC151103 Bos tauru
	7	97.4	7.7	170627	14	AC125567	AC125567 Rattus no
	8	96	7.6	3683	6	AX598999	AX598999 Sequence
	9	94	7.4	93791	14	AC138073	AC138073 Homo sapi
	10	92.4	7.3	6292	6	AX251493	AX251493 Sequence
	11	91.8	7.3	3683	6	AX598853	AX598853 Sequence
c	12	91	7.2	154625	14	CR936360	CR936360 Homo sapi
	13	90.6	7.2	66993	14	AC138074	AC138074 Homo sapi
	14	90.6	7.2	282289	14	AC152242	AC152242 Bos tauru
	15	90.2	7.1	6668	6	AX346599	AX346599 Sequence
	16	88.8	7.0	205309	14	CR847836	CR847836 Danio rer
	17	87.4	6.9	1142	15	AJ840525	AJ840525 Arabidops
	18	86.6	6.8	235711	14	BX908747	BX908747 Danio rer
c	19	86.2	6.8	131682	8	AL672277	AL672277 Human DNA
c	20	85.4	6.7	176162	5	AL953898	AL953898 Zebrafish
c	21	83.6	6.6	76568	2	MBREV	AF538053 Monosiga
c	22	83.2	6.6	197499	5	BX950214	BX950214 Zebrafish
c	23	82.8	6.5	83435	5	BX640469	BX640469 Zebrafish
	24	82.8	6.5	234923	14	AC152762	AC152762 Bos tauru
	25	82.2	6.5	7442	6	AX251441	AX251441 Sequence
c	26	82.2	6.5	246611	14	AC111404	AC111404 Rattus no
c	27	82	6.5	250237	14	AC160076	AC160076 Bos tauru
c	28	82	6.5	289973	14	AC135678	AC135678 Rattus no
	29	80.8	6.4	14867	2	AE001398	AE001398 Plasmodiu
	30	80.6	6.4	1197	15	AJ840663	AJ840663 Arabidops
	31	80.6	6.4	99003	14	AL390756	AL390756 Homo sapi
	32	80.6	6.4	160667	14	AC151623	AC151623 Bos tauru
	33	80.6	6.4	349980	6	AX344560	AX344560 Sequence
	34	80.4	6.4	6314	6	AX281272	AX281272 Sequence
	35	79.8	6.3	212896	14	CR405717	CR405717 Danio rer
	36	79.8	6.3	214213	5	BX629346	BX629346 Zebrafish
c	37	79.6	6.3	84563	5	AC096885	AC096885 Danio rer
c	38	79.6	6.3	212999	14	AC151201	AC151201 Bos tauru
c	39	78.6	6.2	39352	15	AC149394	AC149394 Phakopsor
c	40	78.6	6.2	87140	5	BX571763	BX571763 Zebrafish
c	41	78.6	6.2	103664	5	BX324201	BX324201 Zebrafish
	42	78.2	6.2	286208	14	AC117140	AC117140 Rattus no
	43	78	6.2	14006	6	AX346860	AX346860 Sequence
c	44	78	6.2	110000	14	TANN3_01	Continuation (2 of
c	45	77.4	6.1	1039	6	AR619884	AR619884 Sequence

ALIGNMENTS

RESULT 1
AF078736
LOCUS AF078736 14516 bp DNA linear BCT 03-AUG-1999
DEFINITION Escherichia coli O111 O antigen gene cluster, partial sequence.
ACCESSION AF078736 U13629
VERSION AF078736.1 GI:5678713
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 14516)
AUTHORS Bastin,D.A. and Reeves,P.R.
TITLE Sequence and analysis of the O antigen gene (rfb) cluster of Escherichia coli O111
JOURNAL Gene 164 (1), 17-23 (1995)
PUBMED 7590310
REFERENCE 2 (bases 3021 to 9981)
AUTHORS Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C., Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
TITLE Bacterial polysaccharide synthesis and gene nomenclature
JOURNAL Trends Microbiol. 4 (12), 495-503 (1996)
PUBMED 9004408
REFERENCE 3 (bases 1 to 14516)
AUTHORS Wang,L., Curd,H., Qu,W. and Reeves,P.R.
TITLE Sequencing of Escherichia coli O111 O-antigen gene cluster and identification of O111-specific genes
JOURNAL J. Clin. Microbiol. 36 (11), 3182-3187 (1998)
PUBMED 9774562
REFERENCE 4 (bases 3021 to 9981)
AUTHORS Bastin,D.A.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-1997) Microbiology, University of Sydney, Maze Crescent, Sydney, NSW 2006, Australia
REFERENCE 5 (bases 1 to 14516)
AUTHORS Wang,L., Curd,H., Qu,W., Bastin,D.A. and Reeves,P.R.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-1998) Microbiology, University of Sydney, Maze Crescent, Sydney, NSW 2006, Australia
COMMENT On Aug 3, 1999 this sequence version replaced gi:1155016.
FEATURES Location/Qualifiers

```
source          1. .14516
                /organism="Escherichia coli"
                /mol_type="genomic DNA"
                /strain="Stoke W; M92"
                /serotype="O111"
                /db_xref="taxon:562"
gene            <1. .273
                /gene="galF"
CDS             <1. .273
                /gene="galF"
                /codon_start=1
                /transl_table=11
                /product="putative glucose-1-phosphate
                uridylyltransferase"
                /protein_id="AAD46727.1"
                /db_xref="GI:5678714"
                /translation="DLMAVGRYVLSADIWAELEKTAPGAWGRIQLTDAIAELAKKQSV
                DAMLMTGDSYDCGKKMGYMQAFVKYGLRNLKEGAKFRKSIKKLLSE"
gene            739. .1932
                /gene="wbdH"
CDS             739. .1932
                /gene="wbdH"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase"
                /protein_id="AAD46728.1"
                /db_xref="GI:5678715"
                /translation="MLLCCIHINVYYLLLECDMKKIVIIGNVASMMLRFRKELIMNLV
                RQGDNVYCLANDFSTEDLKVLSSWGVKGVKFSLNSKGINPFKDIIAVYELKKILKDIS
                PDIVFSYFVKPVIFGTIASKLSKVPRIVGMIEGLGNAFTYYKGKQTTKTKMIKWIQIL
                LYKLALPMLDDLILLNHDDKKDLIDQYNIKAKVTVLGGIGLDLNEFSYKEPPKEKITF
                IFIARLLREKGIFEFIEAAKFVKTTYPSSEFVILGGFESNNPFSLQKNEIESLRKEHD
                LIYPGHVENVQDWLEKSSVFVLPTSYREGVPRVIQEAMAIGRPVITTNVPGCRDIIND
                GVNGFLIPPFEINLLAEKMKYFIENKDKVLEMGLAGRKFAEKNFDAFEKNNRLASIIK
                SNNDF"
gene            1983. .3101
                /gene="gmd"
CDS             1983. .3101
                /gene="gmd"
                /codon_start=1
                /transl_table=11
                /product="GDP-D-mannose dehydratase"
                /protein_id="AAD46729.1"
                /db_xref="GI:5678716"
                /translation="MNKVALITGITGQDGSYLAELLLEKGYEVHGIKRRASSFNTERV
                DHIYQDSHLANPKLFLHYGDLTDTSNLTRILKEVQPDEVYNLGAMSHVAVSFESPEYT
                ADVDAIGTLRLLEAIRILGLEKKTKFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
                KLYAYWITVNYRESYGMFACNGILFNHESPRRGETFVTRKITRGIANIAQGLDKCLYL
                GNMDSLRDWGHAKDYVKMQWMMLQQETPEDFVIATGIQYSVREFVTMAAEQVGIELAF
                EGEGVNEKGVVVSVNGTDAKAVNPGDVIISVDPRYFRPAEVETLLGDPTNAHKKLGWS
                PEITLREMVKEMVSSDLAIAKKNVLLKANNIATNIPQE"
gene            3174. .3623
                /gene="wbdI"
CDS             3174. .3623
                /gene="wbdI"
                /note="previously known as orf3.4"
                /codon_start=1
                /transl_table=11
                /product="WbdI"
                /protein_id="AAC44881.1"
                /db_xref="GI:1155017"
                /translation="MFITSDKFREIIKLVPLVSIDLLIENENGEYLFGLRNNRPAKNY
                FFVPGGRIRKNESIKNAFKRISSMELGKEYGISGSVFNGVWEHFYDDGFFSEGEATHY
                IVLCYTLKVLKSELNLPDDQHREYLWLTKHQINAKQDVHNYSKNYFL"
gene            3657. .5108
                /gene="manC"
CDS             3657. .5108
                /gene="manC"
                /EC_number="2.7.7.22"
                /note="previously known as RfbM"
                /codon_start=1
                /transl_table=11
                /product="mannose-1-phosphate guanylyltransferase"
                /protein_id="AAC44882.1"
                /db_xref="GI:1155018"
                /translation="MSQCLYPVIIAGGTGSRLWPLSRVLYPKQFLNLVGDSTMLQTTI
                TRLDGIECENPIVICNEDHRFIVAEQLRQIGKLTKNIILEPKGRNTAPAIALAAFIAQ
                KNNPNDDPLLLVLAADHSINNEKAFRESIIKAMPYATSGKLVTFGIIPDTANTGYGYI
                KRSSSADPNKEFPAYNVAEFVEKPDVKTAQEYISSGNYYWNSGMFLFRASKYLDELRK
                FRPDIYHSCECATATANIDMDFVRINEAEFINCPEESIDYAVMEKTKDAVVLPIDIGW
                NDVGSWSSLWDISQKDCHGNVCHGDVLNHDGENSFIYSESSLVATVGVSNLVIVQTKD
                AVLVADRDKVQNVKNIVDDLKKRKRAEYYMHRAVFRPWGKFDAIDQGDRYRVKKIIVK
                PGEGLDLRMHHHRAEHWIVVSGTAKVSLGSEVKLLVSNESIYIPQGAKYSLENPGVIP
                LHLIEVSSGDYLESDDIVRFTDRYNSKQFLKRD"
gene            5114. .6484
                /gene="manB"
CDS             5114. .6484
                /gene="manB"
                /EC_number="5.4.2.8"
                /note="previously known as RfbK"
                /codon_start=1
                /transl_table=11
                /product="phosphomannomutase"
                /protein_id="AAC44883.1"
                /db_xref="GI:1155019"
                /translation="MNKITCFKAYDIRGRLGAELNDEIAYRIGRAYGEFFKPQTVVVG
                GDARLTSESLKKSLSNGLCDAGVNVLDLGMCGTEEIYFSTWYLGIDGGIEVTASHNPI
                DYNGMKLVTKGARPISSDTGLKDIQQLVESNNFEELNLEKKGNITKYSTRDAYINHLM
                GYANLQKIKKIKIVVNSGNGAAGPVIDAIEECFLRNNIPIQFVKINNTPDGNFPHGIP
                NPLLPECREDTSSAVIRHSADFGIAFDGDFDRCFFFDENGQFIEGYYIVGLLAEVFLG
                KYPNAKIIHDPRLIWNTIDIVESHGGIPIMTKTGHAYIKQRMREEDAVYGGEMSAHHY
                FKDFAYCDSGMIPWILICELLSLTNKKLGELVCGCINDWPASGEINCTLDNPQNEIDK
                LFNRYKDSALAVDYTDGLTMEFSDWRFNVRCSNTEPVVRLNVESRNNAILMQEKTEEI
                LNFISK"
gene            6521. .7444
                /gene="wbdJ"
CDS             6521. .7444
                /gene="wbdJ"
                /note="previously known as orf6.7"
                /codon_start=1
                /transl_table=11
                /product="WbdJ"
                /protein_id="AAC44884.1"
                /db_xref="GI:1155020"
                /translation="MKVLLTGSTGMVGKNILEHDSASKYNILTPTSSDLNLLDKNEIE
                KFMLINMPDCIIHAAGLVGGIHANISRPFDFLEKNLQMGLNLVSVAKKLGIKKVLNLG
                SSCMYPKNFEEAIPEKALLTGELEETNEGYAIAKIAVAKACEYISRENSNYFYKTIIP
                CNLYGKYDKFDDNSSHMIPAVIKKIHHAKINNVPEIEIWGDGNSRREFMYAEDLADLI
                FYVIPKIEFMPNMVNAGLGYDYSINDYYKIIAEEIGYTGSFSHDLTKPTGMKRKLVDI
                SLLNKIGWSSHFELRDGIRKTYNYYLENQNK"
gene            7441. .8607
                /gene="wbdK"
CDS             7441. .8607
                /gene="wbdK"
                /note="previously known as orf7.7"
                /codon_start=1
                /transl_table=11
                /product="WbdK"
                /protein_id="AAC44885.1"
                /db_xref="GI:1155021"
                /translation="MITYPLASNTWDEYEYAAIQSVIDSKMFTMGKKVELYEKNFADL
                FGSKYAVMVSSGSTANLLMIAALFFTNKPKLKRGDEIIVPAVSWSTTYYPLQQYGLKV
                KFVDINKETLNIDIDSLKNAISDKTKAILTVNLLGNPNDFAKINEIINNRDIILLEDN
                CESMGAVFQNKQAGTFGVMGTFSSFYSHHIATMEGGCVVTDDEELYHVLLCLRAHGWT
                RNLPKENMVTGTKSDDIFEESFKFVLPGYNVRPLEMSGAIGIEQLKKLPGFISTRRSN
                AQYFVDKFKDHPFLDIQKEVGESSWFGFSFVIKEGAAIERKSLVNNLISAGIECRPIV
                TGNFLKNERVLSYFDYSVHDTVANAEYIDKNGFFVGNHQIPLFNEIDYLRKVLK"
gene            8646. .9911
                /gene="wzx"
CDS             8646. .9911
                /gene="wzx"
                /note="integral membrane protein; previously known as
                orf8.9"
                /codon_start=1
                /transl_table=11
                /product="putative O unit flippase"
```

```
                     /protein_id="AAC44886.1"
                     /db_xref="GI:1155022"
                     /translation="MVLTVKKILAFGYSKVLPPVIEQFVNPICIFIITPLILNHLGKQ
                     SYGNWILLITIVSFSQLICGGCSAWIAKIIAEQRILSDLSKKNALRQISYNFSIVIIA
                     FAVLISFLILSICFFDVARNNSSFLFAIIICGFFQEVDNLFSGALKGFEKFNVSCFFE
                     VITRVLWASIVIYGIYGNALLYFTCLAFTIKGMLKYILVCLNITGCFINPNFNRVGIV
                     NLLNESKWMFLQLTGGVSLSLFDRLVIPLILSVSKLASYVPCLQLAQLMFTLSASANQ
                     ILLPMFARMKASNTFPSNCFFKILLVSLISVLPCLALFFFGRDILSIWINPTFATENY
                     KLMQILAISYILLSMMTSFHFLLLGIGKSKLVANLNLVAGLALAASTLIAAHYGLYAI
                     SMVKIIYPAFQFYYLYVAFVYFNRAKNVY"
     gene            9901. .10953
                     /gene="wzy"
     CDS             9901. .10953
                     /gene="wzy"
                     /codon_start=1

  Query Match             100.0%;  Score 1266;  DB 1;  Length 14516;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-172;
  Matches 1266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8646 ATGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTT 8705

Qy         61 ATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8706 ATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCAC 8765

Qy        121 CTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8766 CTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAG 8825

Qy        181 TTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8826 TTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTT 8885

Qy        241 AGTGATTTATCAAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8886 AGTGATTTATCAAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATT 8945

Qy        301 ATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8946 ATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCG 9005

Qy        361 AGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9006 AGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGAT 9065

Qy        421 AATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9066 AATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAA 9125

Qy        481 GTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9126 GTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTC 9185

Qy        541 TTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9186 TTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGT 9245

Qy        601 CTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9246 CTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTG 9305

Qy        661 TTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9306 TTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTT 9365

Qy        721 GATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9366 GATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGC 9425

Qy        781 CTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9426 CTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCA 9485

Qy        841 ATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9486 ATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTG 9545

Qy        901 CTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9546 CTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATA 9605

Qy        961 TTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9606 TTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTA 9665

Qy       1021 GCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9666 GCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 9725

Qy       1081 GGTAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9726 GGTAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCA 9785

Qy       1141 ACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9786 ACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCG 9845

Qy       1201 GCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9846 GCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTC 9905

Qy       1261 TATTGA 1266
              ||||||
Db       9906 TATTGA 9911


RESULT 2
BD085554
LOCUS       BD085554               14516 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085554
VERSION     BD085554.1  GI:22631164
KEYWORDS    JP 2001522248-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 14516)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 1 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Escherichia coli
            PN   JP 2001522248-A/1
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14516
            FT                   /organism='Escherichia coli'.
FEATURES             Location/Qualifiers
     source          1. .14516
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 1266;  DB 6;  Length 14516;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-172;
```

```
  Matches 1266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8646 ATGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTT 8705

Qy         61 ATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8706 ATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCAC 8765

Qy        121 CTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8766 CTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAG 8825

Qy        181 TTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8826 TTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTT 8885

Qy        241 AGTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8886 AGTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATT 8945

Qy        301 ATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8946 ATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCG 9005

Qy        361 AGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9006 AGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGAT 9065

Qy        421 AATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9066 AATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAA 9125

Qy        481 GTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9126 GTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTC 9185

Qy        541 TTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9186 TTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGT 9245

Qy        601 CTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9246 CTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTG 9305

Qy        661 TTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9306 TTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTT 9365

Qy        721 GATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9366 GATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGC 9425

Qy        781 CTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9426 CTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCA 9485

Qy        841 ATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9486 ATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTG 9545

Qy        901 CTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9546 CTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATA 9605

Qy        961 TTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9606 TTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTA 9665

Qy       1021 GCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9666 GCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 9725

Qy       1081 GGTAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9726 GGTAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCA 9785

Qy       1141 ACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9786 ACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCG 9845

Qy       1201 GCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9846 GCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTC 9905

Qy       1261 TATTGA 1266
              ||||||
Db       9906 TATTGA 9911
```

RESULT 3
AX029261

LOCUS	AX029261 14520 bp DNA linear PAT 16-SEP-2000
DEFINITION	Sequence 1 from Patent WO9850531.
ACCESSION	AX029261
VERSION	AX029261.1 GI:10190112
KEYWORDS	.
SOURCE	Escherichia coli
ORGANISM	Escherichia coli Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE	1
AUTHORS	Wang,L. and Reeves,P.R.
TITLE	Nucleic acid molecules specific for bacterial antigens and uses thereof
JOURNAL	Patent: WO 9850531-A 1 12-NOV-1998; UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES	Location/Qualifiers
source	1. .14520 /organism="Escherichia coli" /mol_type="unassigned DNA" /db_xref="taxon:562"
ORIGIN	

```
  Query Match             100.0%;  Score 1266;  DB 6;  Length 14520;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-172;
  Matches 1266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8646 ATGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTT 8705

Qy         61 ATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8706 ATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCAC 8765

Qy        121 CTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8766 CTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAG 8825

Qy        181 TTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8826 TTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTT 8885

Qy        241 AGTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8886 AGTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATT 8945

Qy        301 ATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8946 ATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCG 9005

Qy        361 AGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9006 AGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGAT 9065
```

```
Qy        421 AATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9066 AATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAA 9125

Qy        481 GTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9126 GTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTC 9185

Qy        541 TTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9186 TTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGT 9245

Qy        601 CTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9246 CTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTG 9305

Qy        661 TTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9306 TTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTT 9365

Qy        721 GATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9366 GATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGC 9425

Qy        781 CTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9426 CTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCA 9485

Qy        841 ATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9486 ATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTG 9545

Qy        901 CTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9546 CTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATA 9605

Qy        961 TTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9606 TTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTA 9665

Qy       1021 GCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9666 GCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 9725

Qy       1081 GGTAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9726 GGTAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCA 9785

Qy       1141 ACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9786 ACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCG 9845

Qy       1201 GCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9846 GCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTC 9905

Qy       1261 TATTGA 1266
              ||||||
Db       9906 TATTGA 9911


RESULT 4
AF285969
LOCUS       AF285969               13650 bp    DNA     linear   BCT 02-AUG-2001
DEFINITION  Salmonella enterica O antigen gene cluster, complete sequence.
ACCESSION   AF285969
VERSION     AF285969.1  GI:15076983
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 13650)
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     The E. coli O111 and S. enterica O35 gene clusters: Gene clusters
            encoding the same colitose containing O antigen are highly
            conserved
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 13650)
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2000) Department of Microbiology, The University
            of Sydney, Sydney, NSW 2006, Australia
FEATURES             Location/Qualifiers
     source          1. .13650
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /serovar="Adelaide"
                     /db_xref="taxon:28901"
                     /note="serogroup: O35"
     gene            1321. .2460
                     /gene="wbdH"
     CDS             1321. .2460
                     /gene="wbdH"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glycosyl transferase WbdH"
                     /protein_id="AAK83009.1"
                     /db_xref="GI:15076984"
                     /translation="MKKIVIIGNIASMMLRFRKELIINLVSQGDDVYCLANDFSPEDL
                     KLLSSWGGKGIKYTLNSKGINPFKDILSVIELKNILDKISPDIVFSYFVKPVIFGTIA
                     SKLSRVPRIVGMIEGLGNAFTYYKGKQSIKTRIIKWMQIFLYKLALPMLNDLVLLNKD
                     DKKDLIERYNIKVNVTILGGIGLNLNDFSYKEPLTDKVTFIFIARLLREKGVFEYIEA
                     AKRVKNIYPNVEFVILGGFENNNPFSLQQHEIELLRKEHDLIYPGHVENVQDWLEKSS
                     VFVLPTSYREGVPRVIQEAMAIGRPVITTNVPGCKDIINDGINGFLVPPFESELLVKK
                     MIYFIDNRSKILEMGLSGRLFAEKNFDAMEKNKTLASIIKANHEF"
     gene            2511. .3629
                     /gene="gmd"
     CDS             2511. .3629
                     /gene="gmd"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-D-mannose dehydratase Gmd"
                     /protein_id="AAK83010.1"
                     /db_xref="GI:15076985"
                     /translation="MNKVALITGITGQDGSYLAELLLEKGYEVHGIKRRASSFNTERV
                     DHIYQDIHLNNPKLYLHYGDLTDTSNLTRILQEVQPDEVYNLGAMSHVAVSFESPEYT
                     ADVDALGTLRLLEAIRILGLEKKTKFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
                     KLYAYWITVNYRESYGMFACNGILFNHESPRRGETFVTRKITRGIANIAQGLDKCLYL
                     GNMDSLRDWGHAKDYVKMQWMMLQQEAPEDFVIATGVQYSVRDFVTMAAKQVGIELVF
                     EGEGINEKGIVASVNGNDAKAVKAGDVIISVDPRYFRPAEVETLLGDPTNAHKKLGWS
                     PEITLCEMVKEMVSSDLAIAKKNVLLKTNNIPTNIPQE"
     gene            3698. .4147
                     /gene="gmm"
     CDS             3698. .4147
                     /gene="gmm"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-mannosyl hydrolase Gmm"
                     /protein_id="AAK83011.1"
                     /db_xref="GI:15076986"
                     /translation="MFIPSDKFREVIKLTPLVSIDILIENENGEYLFGLRNNRPAKNQ
                     FFVPGGRIRKSEFIENAFKRISFAELGKEYDISESLFNGIWEHFYEESFFSDADATHY
                     IVLCYKLKVLKNELNLPADQHCEYIWISEDKISNLNNIHKYSKDYFL"
     gene            4181. .5626
                     /gene="manC"
     CDS             4181. .5626
                     /gene="manC"
                     /codon_start=1
                     /transl_table=11
                     /product="mannose-1-phosphate guanylyltransferase ManC"
                     /protein_id="AAK83012.1"
                     /db_xref="GI:15076987"
                     /translation="MSQCLYPVVIAGGTGSRLWPLSRVLYPKQFLNLVGNSTMLQTTI
                     ARLNGIKCENPIVICNEDHRFIVAEQLRQIDKLTKNIILEPKGRNTAPAIALAAFIAQ
```

KKKPKDDPLLLVLAADHSINNEQAFREAIIKAIPYADAGKLVMFGIIPSVAHTGYGYI
KRSVPVDADKDSTAYYVANFVEKPNIQKAQEYIMSEGYYWNSGMFLFRASKYLGELKK
FRPDIYNKCESATATANIDMDFVRINETEFINCPEESIDYAVMEKTKDAVVLPIDVGW
SDVGSWSSLWDISQKDVHGNVCQGDVINHDSEDSFIYSESSLVATVGVSNLVIVQTKD
AVLVADRDKVQNVKNIVDSLKQGGRAEYYMHREVFRPWGKYDAIDQGDRYKVKKIIVK
PGEGLDLRMHHHRAEHWIVVSGTAKVSLGNKIKLLVPNESIYIPQGAEYSLENPGVIP
LHLIEVSSGDYLESDDIVRFTDRYNSKKFLK"

gene 5639. .7015

CDS 5639. .7015
/gene="manB"
/codon_start=1
/transl_table=11
/product="phosphomannomutase ManB"
/protein_id="AAK83013.1"
/db_xref="GI:15076988"
/translation="MNKLTCFKAYDIRGRLGVELDDDIAYRIGRAYGELFKPQTVVVG
GDTRLTSESLKKSLSNGLCDAGVNVLDLGMCGTEEIYFSTWHLGVDGGIEVTASHNPI
DYNGMKLVTKNARPISSDTGLKDIQTLVELNDFEKINLEKTGNITKYSTRDAYIKHLM
GYVNLQKIKKIKIVVNSGNGAAGPVIDDIEKYFLQNNIPVTFVKINNTPDGHFPNGIP
NPLLPECREDTSKAVIAHSADFGIAFDGDFDRCFFFDEKGRFIEGYYIVGLLAEVFLE
KYPNAKIIHDPRLIWNTIDIVQNYGGKPIMTKTGHAYIKQRMREEDAVYGGEMSAHHY
FKDFEYCDSGMIPWILICELLSIRNKNLGELVYNCINAWPASGEINCTLSNPQSEIDK
LFFYYKNSAISVDYTDGLSMEFSDWRFNVRCSNTEPVVRLNVESKSNVLLMQEKTREI
LSFIGKCV"

gene 7040. .7963
/gene="wbdJ"

CDS 7040. .7963
/gene="wbdJ"
/codon_start=1
/transl_table=11
/product="WbdJ"
/protein_id="AAK83014.1"
/db_xref="GI:15076989"
/translation="MKVLLTGSSGMVGKNILEHASAKDYEILTPTSSDLNLLVIDEIE
KFMLANKPECIIHAAGLVGGIHANINRPFDFLANNLLMGLNLVTIAKKLGIKKFLNLG
SSCMYPRNFEEAIPEKALLTGELEETNEGYAIAKVAVAKACEYISRENSNYYYKTIIP
CNLYGKYDKFGDDSSHMIPAVIKKIHHAKANNIPEIEIWGDGNSRREFMYAEDLADLV
FYIIPKIESIPNMVNAGLGYDYTVNDYYKIIAEEIGYTGTFFHDLTKPTGMKRKLVDI
SLLNKIGWSSHFNLRDGIRKTYNYYLENQYK"

gene 7960. .9126
/gene="wbdK"

CDS 7960. .9126
/gene="wbdK"
/codon_start=1
/transl_table=11
/product="WbdK"
/protein_id="AAK83015.1"
/db_xref="GI:15076990"
/translation="MITYPLASNTWDENEYAAIQSVIDSKMFTMGKNVELYEKNFADL
FGSKYAVMVSSGSTANLLMIAALFFTHKPKLKRGDEIIVPAVSWSTTYYPLQQYGLKV
KFVDINKETLNIDVDSLKNAISDKTRAILAVNLLGNPNDFSTIHEIISNKDIILLEDN
CESMGATFQNKQAGTFGIMGSFSSFYSHHIATMEGGCVVTDDEELYHILLCLRAHGWT
RNLPKENMVTGTKSDDIFEESFKFVLPGYNVRPLEMSGAIGIEQLKKLPGFISTRRSN
AQYFVDKFKDHPFLDIQKEIGESSWFGFSLIIKQDTGIERKSLINKLITSGIECRPIV
TGNFLKNKRVLSYFDYSVHDTVVNAEYIDKNGFFVGNHQIPLLNEIDFLRKVLK"

gene 9178. .10431
/gene="wzx"

CDS 9178. .10431
/gene="wzx"
/codon_start=1
/transl_table=11
/product="O antigen flippase Wzx"
/protein_id="AAK83016.1"
/db_xref="GI:15076991"
/translation="MKKFLAFGYSKVLPPIIEQFINPICIFVITPLILNHLGKQSYGN
WILLITIVSFSQLICGGCSAWIAKVIAEQRATSDYAGKGALRKIAYNLSIIMLSFAIL
ILLILLSLCVFDVARDNTEFLYAIAICGFFQEVDNLFNGALKGFERFNVSCLFEVITR
VLWAFIVIYGIQENSVLYYTCLAFILKGTLKYILVCLNISGKFIYPNFDKAGIVNLLN
ESKWMFLQLTGGISLGLFDRLVIPLILSVNKLSSYVPCLQLVQLMFTISASANQILLP
MFARMKASNTFPSNCFFKIQLVSLSSALPCLVLLFFGYDILSIWISPAFASENFKLMQ
ILSISYILLSMMTSFHFLLLGLGKSKLVANLNLIAGLTLAASTLIAAHYGLYAISVVK
IIYPLFQFYYLYVAFFYFKREKNVY"

gene 10421. .11485
/gene="wzy"

CDS 10421. .11485
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O antigen polymerase Wzy"
/protein_id="AAK83017.1"
/db_xref="GI:15076992"
/translation="MSIDFLFSITEIAVVFSCAIYIFVQCLLMHKIYLDKVVLILLCL
LLCLVIIQLPELNEQGLIDAVKLLLPLLIVFIAFQKPKLCLWIIIALLFFNSTLNFWH
FKVFDKFSSFPFTFFILLYFLYKLVNDQPPRQRNKKFYVVIFIFILIDVTQSLLISYR
GQILYSLICVIMLVLRINLRKTIPYIFLTLPFIYVFIMTFIGFNYFNKDVVFFEPTAS
NIERTGMIYYLVIHANDYIFHGMGTLHFLNEGGQYKMLYGFPSLIPNDPHDFLLRFFI
SIGMMGALIYHCIFFVFFKNISSLLRQKNSYFVIVSCLLLLQVVLFYTLNPFNAFNRL
ICGLTIGVIYGFSKSRQIYM"

gene 11451. .12344
/gene="wbdL"

CDS 11451. .12344
/gene="wbdL"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdL"
/protein_id="AAK83018.1"
/db_xref="GI:15076993"
/translation="MDFQRVDRYICSESFNTPLISIIIATFNSELVIANCLQSVISQS
YKNIEVIIMDGGSYDKTLDIARSFKDERIKIVSEKDRGIYDAWNKAVDLSSGDWIAFI
GSDDVYYHTDAITSLVKGAAISNGAPVVYGRTAHEGPNKEISGFSGSEWYNLKGFKFN
YYKCNLPLPIMSAIYSRDFFKDQRFDIKLKIVADADWFLRCFIKWRKENSPYFIDDIT
PVVRMGYGGVSTDISSQIKTTRESFIVRKKNNISCFNMQLILRYIKIMVMVSIKNIFG
NNIYNSMHNGYHFLKKIKNKI"

gene 12341. .13465
/gene="wbdM"

CDS 12341. .13465
/gene="wbdM"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdM"
/protein_id="AAK83019.1"
/db_xref="GI:15076994"
/translation="MKILYVITGLTCGGAEHLMMQLADQMLLRGHDVNIICLSGISEI
QPTQKHNIHYIRMEKNVRGFVKALFQVRKIITVIKPDIIHSHMFHANIFSRIIRILTP
SVPLICTAHNKNEGGYARMLCYRLSDFLASITTNVSQEAVQEFIAKKAAPKNKIIEVP
NFVNTTNFSFDLKARKERRNFFGIKDNTLLLLAVGRLVDAKDYPNLLNAINYLTLSES
TGFDNFILFIAGDGVLKSKLVDLVNKLNLADKVFFLGRRSDIKELMCAADLFVLSSQW

Query Match 71.6%; Score 906.6; DB 1; Length 13650;
Best Local Similarity 82.3%; Pred. No. 3.7e-121;
Matches 1041; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

```
Qy          2 TGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTTA 61
              ||||||||| |||||||||| |||||||||| |||||||||||||| |||||||| |||
Db       9167 TGGTATTAATAGTGAAAAAATTTTTAGCGTTCGGCTATTCTAAAGTATTACCACCGATTA 9226

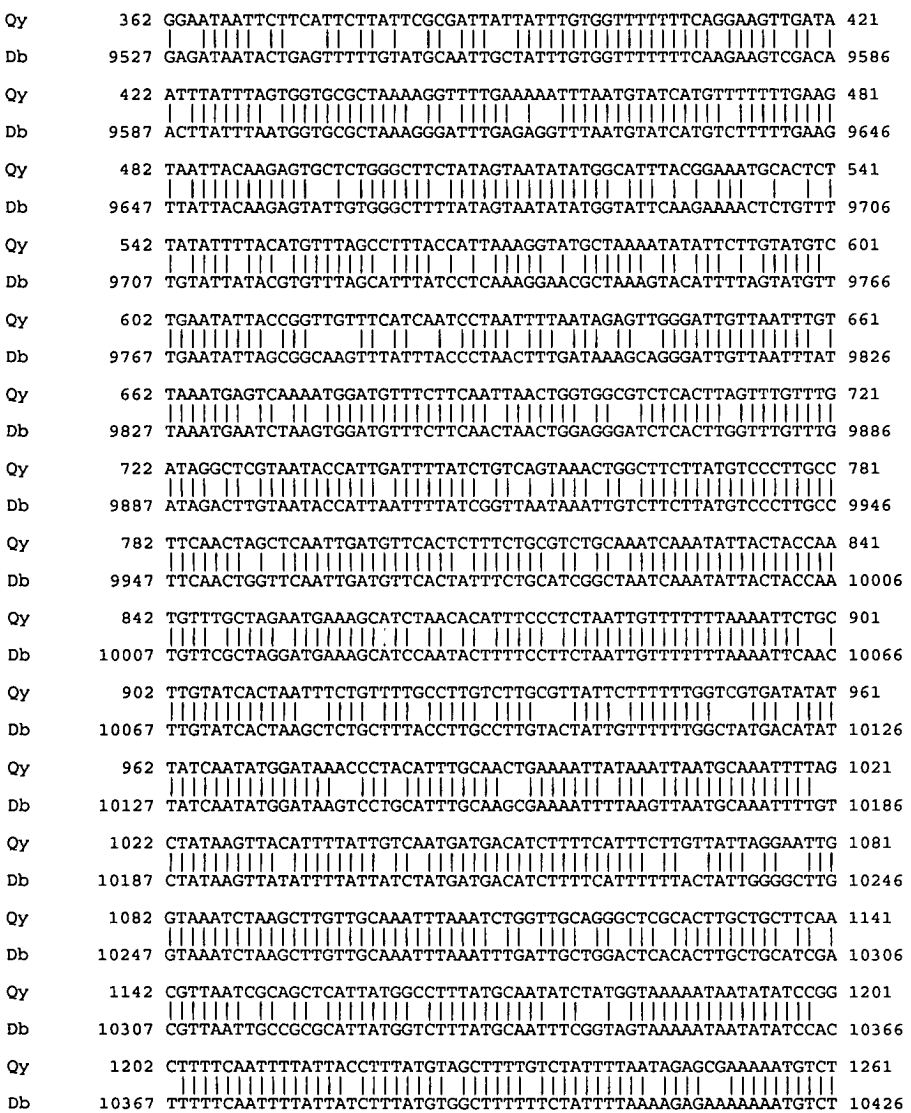
Qy         62 TTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCACC 121
              ||||||| ||| | |||||||||||| || ||| | || || ||| |||||||||| ||
Db       9227 TTGAACAATTTATTAATCCAATTTGTATTTTCGTGATTACGCCATTAATACTCAATCATT 9286

Qy        122 TGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGT 181
              ||||||| |||||||||||| |||||||||||||||||| || || ||||| |||||||||
Db       9287 TGGGTAAACAAAGCTATGGCAATTGGATTTTATTAATCACCATCGTATCCTTTTCTCAGC 9346

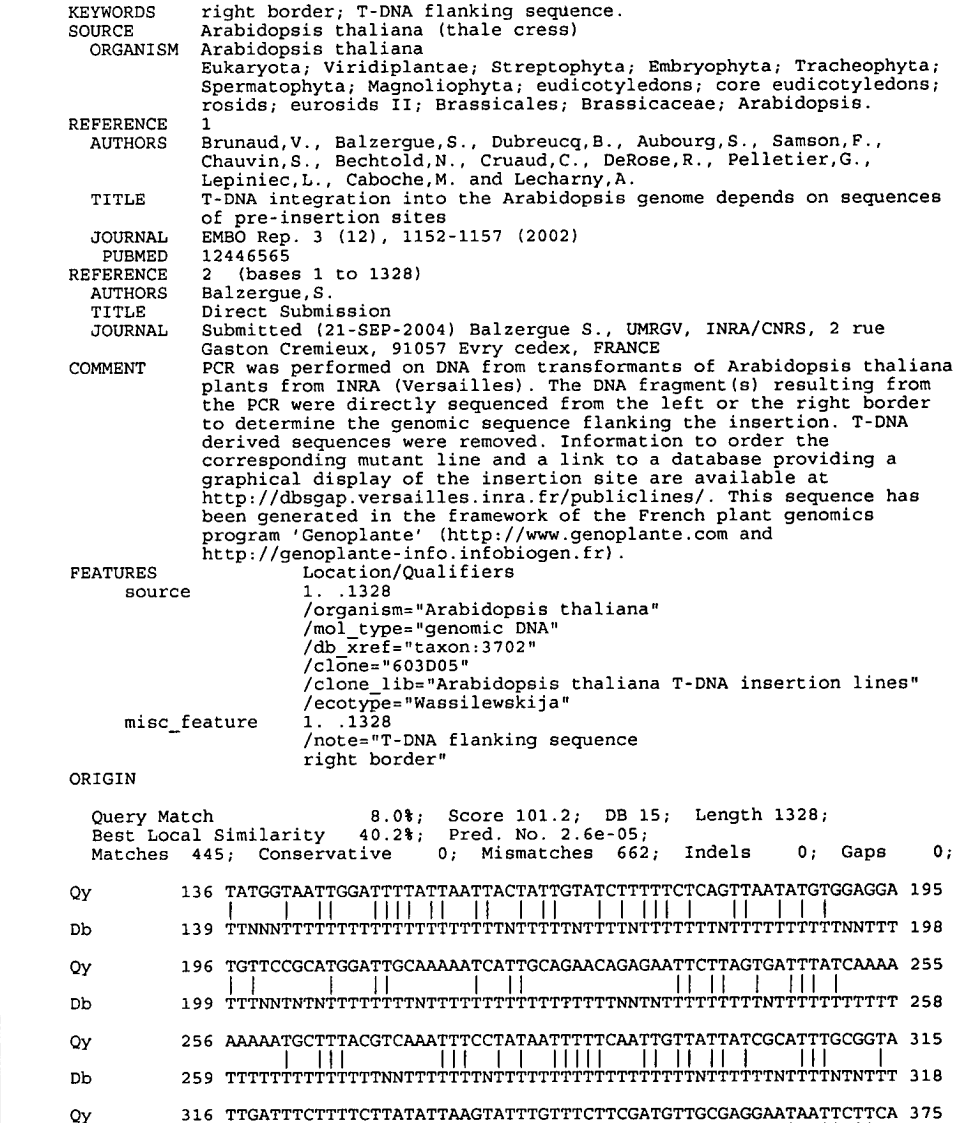
Qy        182 TAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTA 241
              |||||||||| ||||| || || |||||||| ||| | |||||||||||||||    ||
Db       9347 TAATATGTGGCGGATGCTCTGCGTGGATTGCGAAAGTAATTGCAGAACAGAGAGCCACTA 9406

Qy        242 GTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTA 301
              ||||||    |  ||||  ||||||||  ||||  |||||||| ||||||| ||||
Db       9407 GTGATTATGCGGGAAAAGGTGCTTTACGAAAAATAGCCTATAATCTTTCAATTATTATGC 9466

Qy        302 TCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGA 361
              || | ||||| | | |||| || | || |||| |||| | ||||||||||||||
Db       9467 TCTCTTTTGCAATTCTTATTTTATTAATATTACTAAGCCTTTGCGTTTTCGATGTTGCGA 9526
```



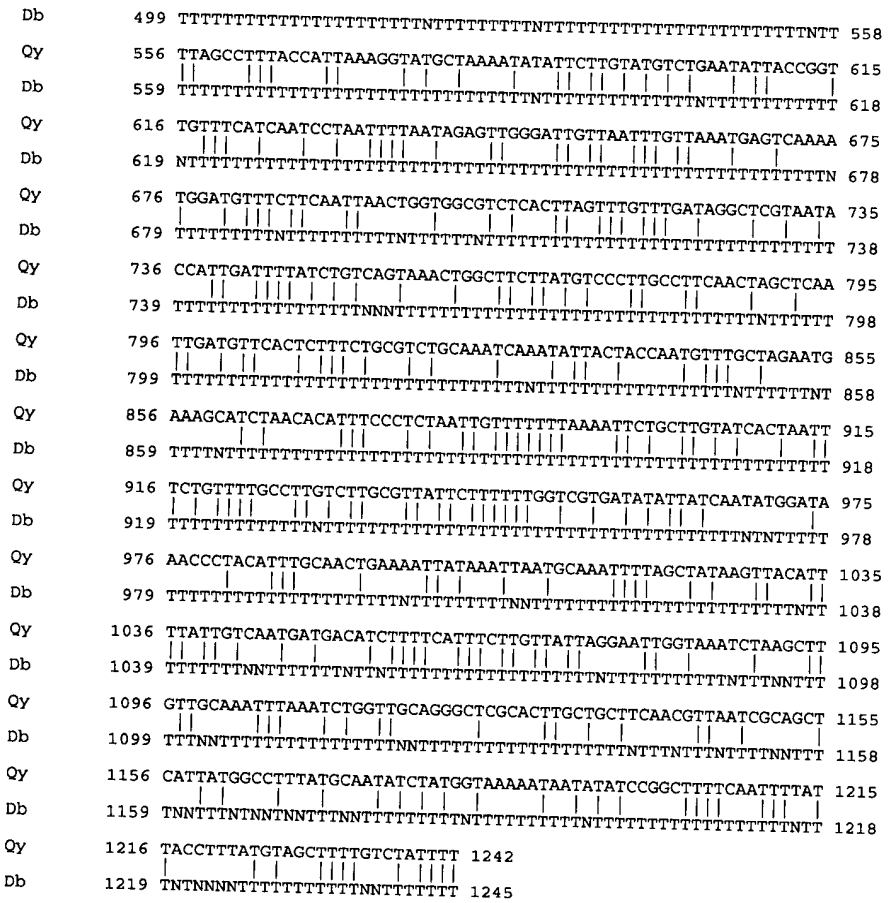
RESULT 5
AJ840631
LOCUS AJ840631 1328 bp DNA linear PLN 22-SEP-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, right border, clone 603D05.
ACCESSION AJ840631
VERSION AJ840631.1 GI:52544837



KEYWORDS right border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F., Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G., Lepiniec,L., Caboche,M. and Lecharny,A.
TITLE T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)
PUBMED 12446565
REFERENCE 2 (bases 1 to 1328)
AUTHORS Balzergue,S.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-2004) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at http://dbsgap.versailles.inra.fr/publiclines/. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (http://www.genoplante.com and http://genoplante-info.infobiogen.fr).
FEATURES Location/Qualifiers
source 1. .1328
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="603D05"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Wassilewskija"
misc_feature 1. .1328
/note="T-DNA flanking sequence
right border"
ORIGIN

Query Match 8.0%; Score 101.2; DB 15; Length 1328;
Best Local Similarity 40.2%; Pred. No. 2.6e-05;
Matches 445; Conservative 0; Mismatches 662; Indels 0; Gaps 0;





RESULT 6
AC151103
LOCUS AC151103 223697 bp DNA linear HTG 01-JUL-2005
DEFINITION Bos taurus clone CH240-4B3, WORKING DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION AC151103
VERSION AC151103.3 GI:68227174
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 223697)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 223697)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (20-AUG-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 223697)
AUTHORS Cow Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jun 26, 2005 this sequence version replaced gi:58038070.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

```
            -------------- Project Information
                Center project name: FAEO
                Center clone name: CH240-4B3
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 218866 bases at least Q40
                Consensus quality: 219341 bases at least Q30
                Consensus quality: 220036 bases at least Q20
                Estimated insert size: 229160; sum-of-contigs estimation
                Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    70653: contig of 70653 bp in length
            *    70654    70706: gap of 53 bp
            *    70707    80327: contig of 9621 bp in length
            *    80328    80377: gap of 50 bp
            *    80378   102971: contig of 22594 bp in length
            *   102972   103021: gap of 50 bp
            *   103022   160686: contig of 57665 bp in length
            *   160687   160802: gap of 116 bp
            *   160803   161700: contig of 898 bp in length
            *   161701   161750: gap of 50 bp
            *   161751   189651: contig of 27901 bp in length
            *   189652   189751: gap of unknown length
            *   189752   191193: contig of 1442 bp in length
            *   191194   191293: gap of unknown length
            *   191294   223697: contig of 32404 bp in length.
FEATURES             Location/Qualifiers
     source          1. .223697
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /clone="CH240-4B3"
     gap             70654. .70706
                     /estimated_length=53
     gap             80328. .80377
                     /estimated_length=50
     gap             102972. .103021
                     /estimated_length=50
     gap             160687. .160802
                     /estimated_length=116
     gap             161701. .161750
                     /estimated_length=50
     gap             189652. .189751
                     /estimated_length=unknown
     gap             191194. .191293
                     /estimated_length=unknown
ORIGIN

  Query Match             7.8%;  Score 98.4;  DB 14;  Length 223697;
  Best Local Similarity   42.2%;  Pred. No. 8.1e-06;
  Matches  354;  Conservative    0;  Mismatches  484;  Indels    0;  Gaps    0;

Qy        236 TTCTTAGTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTG 295
              || ||  |  ||| |           |  |||    |    |||  |    |||||   ||
Db     160857 TTTTTTTTNNTTTTTTTTTNNTTTTTTTTTTTTTTTNNTTTTTTNNTTTTTTTTTTTT 160916

Qy        296 TTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATG 355
              || || |    |||    | ||  ||| |||| || | ||   | ||| ||| ||   |
Db     160917 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 160976

Qy        356 TTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAG 415
              ||      |  || ||  || || ||     || || ||| |  |||||||
Db     160977 TTTTTTTNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 161036

Qy        416 TTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTT 475
              ||  |  ||| |||  |  |  |     |      ||||      |||  | | |    |||||
Db     161037 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTTTTTT 161096

Qy        476 TTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATG 535
              ||     |  ||        |  | |     || | |  |  | |       |||
Db     161097 TTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTNNT 161156

Qy        536 CACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTG 595
                | || | ||||   | |||     |||     ||      | |  |     | | ||  ||
Db     161157 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTTT 161216

Qy        596 TATGTCTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTA 655
              | | | |   | ||    || |||  |    |  |  ||||  |    ||    || ||
Db     161217 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 161276

Qy        656 ATTTGTTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTT 715
               ||| ||   |   |          | ||| ||   ||   |  |    | |   ||  ||
Db     161277 TTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTTTTTT 161336

Qy        716 TGTTTGATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCC 775
              | |||  |    |  |      ||  |||| | | |    |    |   || || | |
Db     161337 TTTTTNNTTNNTTTTTTNNTTTTNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNN 161396

Qy        776 CTTGCCTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTAC 835
               ||   ||    |   |   ||  | ||   | ||      | |     |    | ||
Db     161397 TTTTNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNNNTTTTTTTTTTTTNNTTTTTT 161456

Qy        836 TACCAATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAA 895
              |      | |||  |    |      | |      |||    | |   | |||||||
Db     161457 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNNTTTTTTTTTTT 161516

Qy        896 TTCTGCTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTG 955
              || |  || | |  | ||| | ||||  || | ||   || || ||||||  |  |
Db     161517 TTTTTTTTTTTTNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 161576

Qy        956 ATATATTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAA 1015
              | | || |   | |  |      |  |||      |     || |   ||   |
Db     161577 TTTTTTTTTNNTTTTTTTTNNTTTTTTTTTTTTTTTTTTTTTTTTTNNNTTTTTTTTTT 161636

Qy       1016 TTTTAGCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATT 1073
              ||||   | |   ||   |||| || |   |  |    | ||||  ||| || |  ||
Db     161637 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTT 161694


RESULT 7
AC125567
LOCUS       AC125567              170627 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-9H22, *** SEQUENCING IN PROGRESS ***,
            4 unordered pieces.
ACCESSION   AC125567
VERSION     AC125567.5  GI:24817949
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 170627)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
```

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 170627)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 170627)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 9, 2002 this sequence version replaced gi:23096537.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GDDV
Center clone name: CH230-9H22
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 162619 bases at least Q40
Consensus quality: 164392 bases at least Q30
Consensus quality: 165269 bases at least Q20
Estimated insert size: 169231; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 109158: contig of 109158 bp in length
* 109159 109258: gap of unknown length
* 109259 110280: contig of 1022 bp in length
* 110281 110380: gap of unknown length
* 110381 111540: contig of 1160 bp in length
* 111541 111640: gap of unknown length
* 111641 170627: contig of 58987 bp in length.

FEATURES Location/Qualifiers
source 1. .170627
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-9H22"
misc_feature 1. .1284
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 3636. .4924
/note="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH341988"
gap 109159. .109258
/estimated_length=unknown
gap 110281. .110380
/estimated_length=unknown
gap 111541. .111640
/estimated_length=unknown
ORIGIN

Query Match 7.7%; Score 97.4; DB 14; Length 170627;
Best Local Similarity 45.2%; Pred. No. 1.3e-05;
Matches 447; Conservative 0; Mismatches 531; Indels 11; Gaps 2;

```
Qy      276 TTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATATT 335
            ||  |||| || |  |||| ||||||     ||||        || |||| |||| |||||||
Db    28080 TTTTTATATTTATAAAATTTTTATTAAATTATTTTTTTATTTTATTTTTTTTTTTATATT 28139

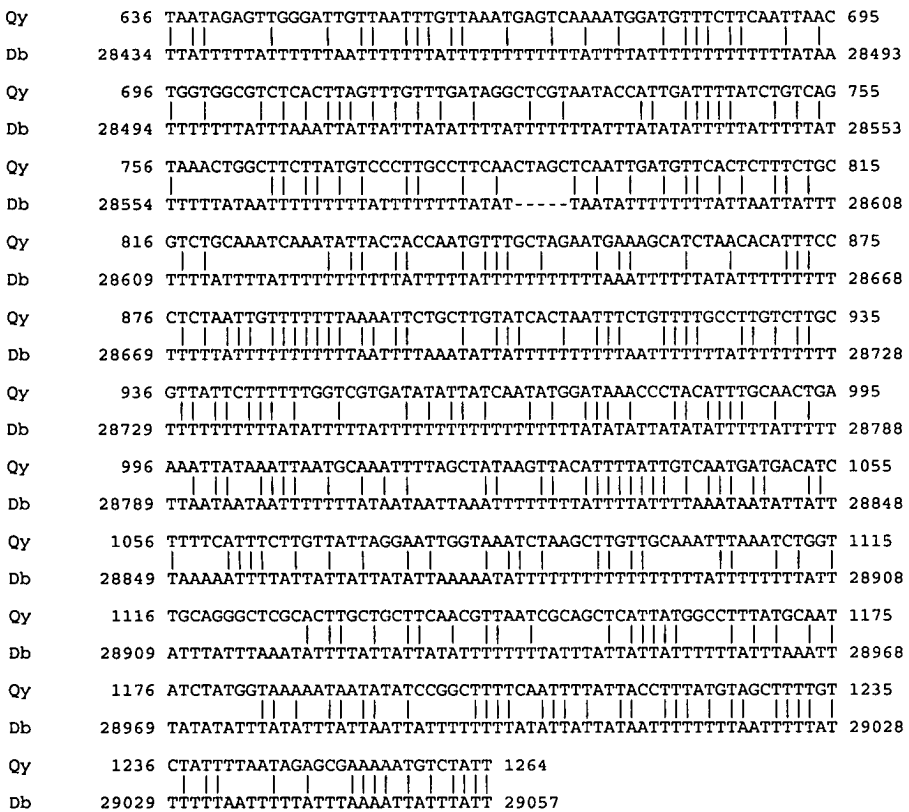
Qy      336 AAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTAT 395
            | ||  | | ||  | ||       | || |  | ||| || ||     || || |
Db    28140 TTTTTTTATTATTTTTTTTTTTTTTTTATTATTAATATATTTTTTTTTTTTTTTTTTT 28199

Qy      396 TTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAA 455
            || |  |||||||        ||  || | |||||  |       | ||| || |  |
Db    28200 TTTTTATTTTTTTT------TTTTTATTATATTTATTTTAATTTTTAAATTTTATTTTAT 28253

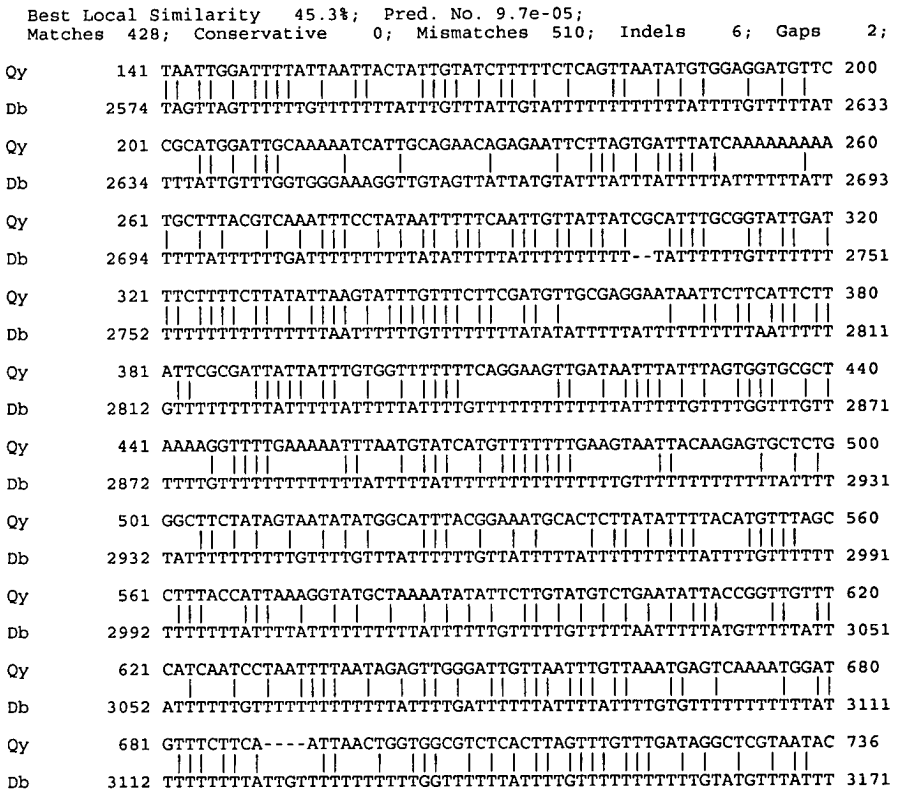
Qy      456 ATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAAT 515
            ||||    | |    |||||||  |   | |      |          || | |  || |
Db    28254 ATTTTTAATTTTTATTTTTTTTTTATATTTAAATTTTATTTATATTTTTTTTTTTTTATT 28313

Qy      516 ATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAGG 575
            || |   | ||          | | || |  |||  || | |    | |    ||| |
Db    28314 ATTTTTTAATTTTTTTTTTTTTATTTTTTTTATTTTTATTTATTTAATATTTATTTATAAT 28373

Qy      576 TATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCATCAATCCTAATTT 635
            |  |||| | | || |  | | | | | ||||      | |||  |   |  ||||
Db    28374 ATTATTAAATTTTTTTTTATTTTTTTTAATTATTTTTTTATTTTTTTTTTTTTTTATTT 28433
```



Query Match 7.6%; Score 96; DB 6; Length 3683;



Best Local Similarity 45.3%; Pred. No. 9.7e-05;
Matches 428; Conservative 0; Mismatches 510; Indels 6; Gaps 2;

RESULT 8
AX598999
LOCUS AX598999 3683 bp DNA linear PAT 14-FEB-2003
DEFINITION Sequence 339 from Patent WO02077272.
ACCESSION AX598999
VERSION AX598999.1 GI:28399139
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Berlin,K., Braun,A., Distler,J., Guetig,D., Howe,A., Mueller,J., Olek,A., Piepenbrock,C., Adorjan,P., Grabs,G., Lesche,R., Leu,E., Lewin,A., Lipscher,E., Maier,S., Model,F., Mueller,V., Otto,T., Pelet,C. and Ziebarth,H.
TITLE Methods and nucleic acids for the analysis of hematopoietic cell proliferative disorders
JOURNAL Patent: WO 02077272-A 339 03-OCT-2002;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
source 1. .3683
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

RESULT 9
AC138073
LOCUS AC138073 93791 bp DNA linear HTG 12-DEC-2002
DEFINITION Homo sapiens chromosome 17 clone RP11-430B7 map 17, LOW-PASS SEQUENCE SAMPLING.

```
ACCESSION   AC138073
VERSION     AC138073.1  GI:26553380
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 93791)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-430B7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 93791)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L28756
                Center clone name: 430_B_7
            --------------.
            * NOTE: This record contains 68 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1     1288: contig of 1288 bp in length
            *     1289     1388: gap of 100 bp
            *     1389     2656: contig of 1268 bp in length
            *     2657     2756: gap of 100 bp
            *     2757     4013: contig of 1257 bp in length
            *     4014     4113: gap of 100 bp
            *     4114     5362: contig of 1249 bp in length
            *     5363     5462: gap of 100 bp
            *     5463     6748: contig of 1286 bp in length
            *     6749     6848: gap of 100 bp
            *     6849     8101: contig of 1253 bp in length
            *     8102     8201: gap of 100 bp
            *     8202     9503: contig of 1302 bp in length
            *     9504     9603: gap of 100 bp
            *     9604    10826: contig of 1223 bp in length
            *    10827    10926: gap of 100 bp
            *    10927    12171: contig of 1245 bp in length
            *    12172    12271: gap of 100 bp
            *    12272    13557: contig of 1286 bp in length
            *    13558    13657: gap of 100 bp
            *    13658    14957: contig of 1300 bp in length
            *    14958    15057: gap of 100 bp
            *    15058    16339: contig of 1282 bp in length
            *    16340    16439: gap of 100 bp
            *    16440    17747: contig of 1308 bp in length
            *    17748    17847: gap of 100 bp
            *    17848    19125: contig of 1278 bp in length
            *    19126    19225: gap of 100 bp
            *    19226    20543: contig of 1318 bp in length
            *    20544    20643: gap of 100 bp
            *    20644    21933: contig of 1290 bp in length
            *    21934    22033: gap of 100 bp
            *    22034    23288: contig of 1255 bp in length
            *    23289    23388: gap of 100 bp
            *    23389    24637: contig of 1249 bp in length
            *    24638    24737: gap of 100 bp
            *    24738    26080: contig of 1343 bp in length
            *    26081    26180: gap of 100 bp
            *    26181    27463: contig of 1283 bp in length
            *    27464    27563: gap of 100 bp
            *    27564    28816: contig of 1253 bp in length
            *    28817    28916: gap of 100 bp
            *    28917    30228: contig of 1312 bp in length
            *    30229    30328: gap of 100 bp
            *    30329    31624: contig of 1296 bp in length
            *    31625    31724: gap of 100 bp
            *    31725    33059: contig of 1335 bp in length
            *    33060    33159: gap of 100 bp
            *    33160    34450: contig of 1291 bp in length
            *    34451    34550: gap of 100 bp
            *    34551    35813: contig of 1263 bp in length
            *    35814    35913: gap of 100 bp
            *    35914    37183: contig of 1270 bp in length
            *    37184    37283: gap of 100 bp
            *    37284    38505: contig of 1222 bp in length
            *    38506    38605: gap of 100 bp
            *    38606    39921: contig of 1316 bp in length
            *    39922    40021: gap of 100 bp
            *    40022    41265: contig of 1244 bp in length
            *    41266    41365: gap of 100 bp
            *    41366    42680: contig of 1315 bp in length
            *    42681    42780: gap of 100 bp
            *    42781    44112: contig of 1332 bp in length
            *    44113    44212: gap of 100 bp
            *    44213    45557: contig of 1345 bp in length
            *    45558    45657: gap of 100 bp
            *    45658    47001: contig of 1344 bp in length
            *    47002    47101: gap of 100 bp
            *    47102    48400: contig of 1299 bp in length
            *    48401    48500: gap of 100 bp
            *    48501    49780: contig of 1280 bp in length
            *    49781    49880: gap of 100 bp
            *    49881    51183: contig of 1303 bp in length
            *    51184    51283: gap of 100 bp
            *    51284    52519: contig of 1236 bp in length
            *    52520    52619: gap of 100 bp
            *    52620    53880: contig of 1261 bp in length
            *    53881    53980: gap of 100 bp
            *    53981    55274: contig of 1294 bp in length
            *    55275    55374: gap of 100 bp
            *    55375    56691: contig of 1317 bp in length
            *    56692    56791: gap of 100 bp
            *    56792    58065: contig of 1274 bp in length
            *    58066    58165: gap of 100 bp
            *    58166    59539: contig of 1374 bp in length
            *    59540    59639: gap of 100 bp
            *    59640    60905: contig of 1266 bp in length
            *    60906    61005: gap of 100 bp
            *    61006    62302: contig of 1297 bp in length
```

```
            *    62303    62402: gap of 100 bp
            *    62403    63596: contig of 1194 bp in length
            *    63597    63696: gap of 100 bp
            *    63697    64984: contig of 1288 bp in length
            *    64985    65084: gap of 100 bp
            *    65085    66268: contig of 1184 bp in length
            *    66269    66368: gap of 100 bp
            *    66369    67636: contig of 1268 bp in length
            *    67637    67736: gap of 100 bp
            *    67737    69026: contig of 1290 bp in length
            *    69027    69126: gap of 100 bp
            *    69127    70399: contig of 1273 bp in length
            *    70400    70499: gap of 100 bp
            *    70500    71763: contig of 1264 bp in length
            *    71764    71863: gap of 100 bp
            *    71864    73116: contig of 1253 bp in length
            *    73117    73216: gap of 100 bp
            *    73217    74715: contig of 1499 bp in length
            *    74716    74815: gap of 100 bp
            *    74816    76106: contig of 1291 bp in length
            *    76107    76206: gap of 100 bp
            *    76207    77422: contig of 1216 bp in length
            *    77423    77522: gap of 100 bp
            *    77523    78820: contig of 1298 bp in length
            *    78821    78920: gap of 100 bp
            *    78921    80230: contig of 1310 bp in length
            *    80231    80330: gap of 100 bp
            *    80331    81610: contig of 1280 bp in length
            *    81611    81710: gap of 100 bp
            *    81711    83067: contig of 1357 bp in length
            *    83068    83167: gap of 100 bp
            *    83168    84434: contig of 1267 bp in length
            *    84435    84534: gap of 100 bp
            *    84535    85791: contig of 1257 bp in length
            *    85792    85891: gap of 100 bp
            *    85892    86978: contig of 1087 bp in length
            *    86979    87078: gap of 100 bp
            *    87079    88352: contig of 1274 bp in length
            *    88353    88452: gap of 100 bp
            *    88453    89751: contig of 1299 bp in length
            *    89752    89851: gap of 100 bp
            *    89852    91085: contig of 1234 bp in length
            *    91086    91185: gap of 100 bp
            *    91186    92443: contig of 1258 bp in length
            *    92444    92543: gap of 100 bp
            *    92544    93791: contig of 1248 bp in length.
FEATURES             Location/Qualifiers
     source          1. .93791
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"

  Query Match             7.4%;  Score 94;  DB 14;  Length 93791;
  Best Local Similarity   32.3%;  Pred. No. 5e-05;
  Matches  310;  Conservative    0;  Mismatches  649;  Indels    0;  Gaps    0;

Qy        264 TTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTC 323
              |||   |    |||  | |  |||||       ||  | |    |||    | |   |||
Db      48821 TTTTTTTNTTTTTTTTTTTNTTTTTTTTTNNNTTNNTTTTTTTTTTTNTTTTNNTTTTT 48880

Qy        324 TTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATT 383
              ||||  ||  | |     | ||  |||  ||    |  |          |  |  ||  || || |
Db      48881 TTTTTTTTTTTNNTTNTTNTTTTTTTTTTTNTTTTNNNTNTTTNTTTTTTTTTTTTTN 48940

Qy        384 CGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAA 443
                || ||  |||  |  |||||  |        ||   |  |||  |||   |   |     |
Db      48941 TTTTTTTNTTTTTTTTNNTTTTTNTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 49000

Qy        444 AGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGC 503
                ||||        ||  | | |  |  || |||        ||        |  | |
Db      49001 TTNTTTTNNTTTTNTTTTTTTNTTTTNNTTNTTTNTNTNTTTTTNTTTTTTTTTTTTNN 49060

Qy        504 TTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTT 563
```

```
              || | |  |  | | |    |          |    | |  | ||||   | |||    |
Db      49061 TTTTTTNTTNTTTTNTTTTTTNNNNNTTTTTTTNTTNTNTTNTTTTTTTTTTTTTTNTTN 49120

Qy        564 TACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCAT 623
              |      |  |       | | || || | | |  |    |  |  ||| |
Db      49121 TNNTTTNTTTTTTTTTNNTTTTTTTTTTTTTNTTTTTNTNTTTTNTNTTTTNTTTTNT 49180

Qy        624 CAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTT 683
                |  |  ||||  |    ||    |   |  ||| ||   |   |     |   | ||
Db      49181 TTTTTTTTTTTTTNTTNTNTTTTTTNTNNNTNTTTTTTTTTTTTTTTNNNNNTTTNTNTT 49240

Qy        684 TCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGAT 743
              | ||   ||       |    | |   ||  ||  | |  |       |  |    ||
Db      49241 TTTTNNTTTTTTNTTTNNTTTTTNTTTTTTTTTNTTNTTTTNNNTNNNTNTTTTTNTTTNN 49300

Qy        744 TTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATGTT 803
              | | |   |    |  || || | |   ||   ||       |       ||    |
Db      49301 NNNTTTTTTTTTTTNTTTTTTTTTTTTTTTTNNNTTNNNTTNTNTTTTTNNTTTTTTNTTN 49360

Qy        804 CACTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCATC 863
              | | |   | |      |          |     | |||
Db      49361 NNTNTTNTTTNTTTTTTTTTTNNNTNNNNTTTNTTTTTTTTTNNNNNNNNNNNNTTNNN 49420

Qy        864 TAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTTTT 923
                  |     |      |||||||     |    || |         ||  | |
Db      49421 NNNNNNTNNTTTTNNTTNNNNTTTTTTTTTTTNTNNTTTTTTNNTTNNNNTTNNTTTNNN 49480

Qy        924 GCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATAAACCCTAC 983
              || | ||   ||  | || |||  | |     |    |  | |  |
Db      49481 NTTTTTTTTTTTNTTTNTTTTNTTTTTTTTNTTNNTNNNNTTTNTTTTTNTTNNNNNNNN 49540

Qy        984 ATTTGCAACTGAAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTTTATTGTC 1043
              |||     |     || |     |    ||||        |     |||  |
Db      49541 NTTTTTTTTTTTTTTTTTTTTNNTTTTNNNTTTTNNNNTTTTNTNNNTTTTNTNNT 49600

Qy       1044 AATGATGACATCTTTTCATTTCTTGTTATTAGGAATTGGTAAATCTAAGCTTGTTGCAAA 1103
              | |     | |                ||          |   |          |
Db      49601 NTTNTTTTNTTNTNNNNNNNNNNNNNNNTTTTTNNNNTTTNTNTNNNNNNNNNTNNNTT 49660

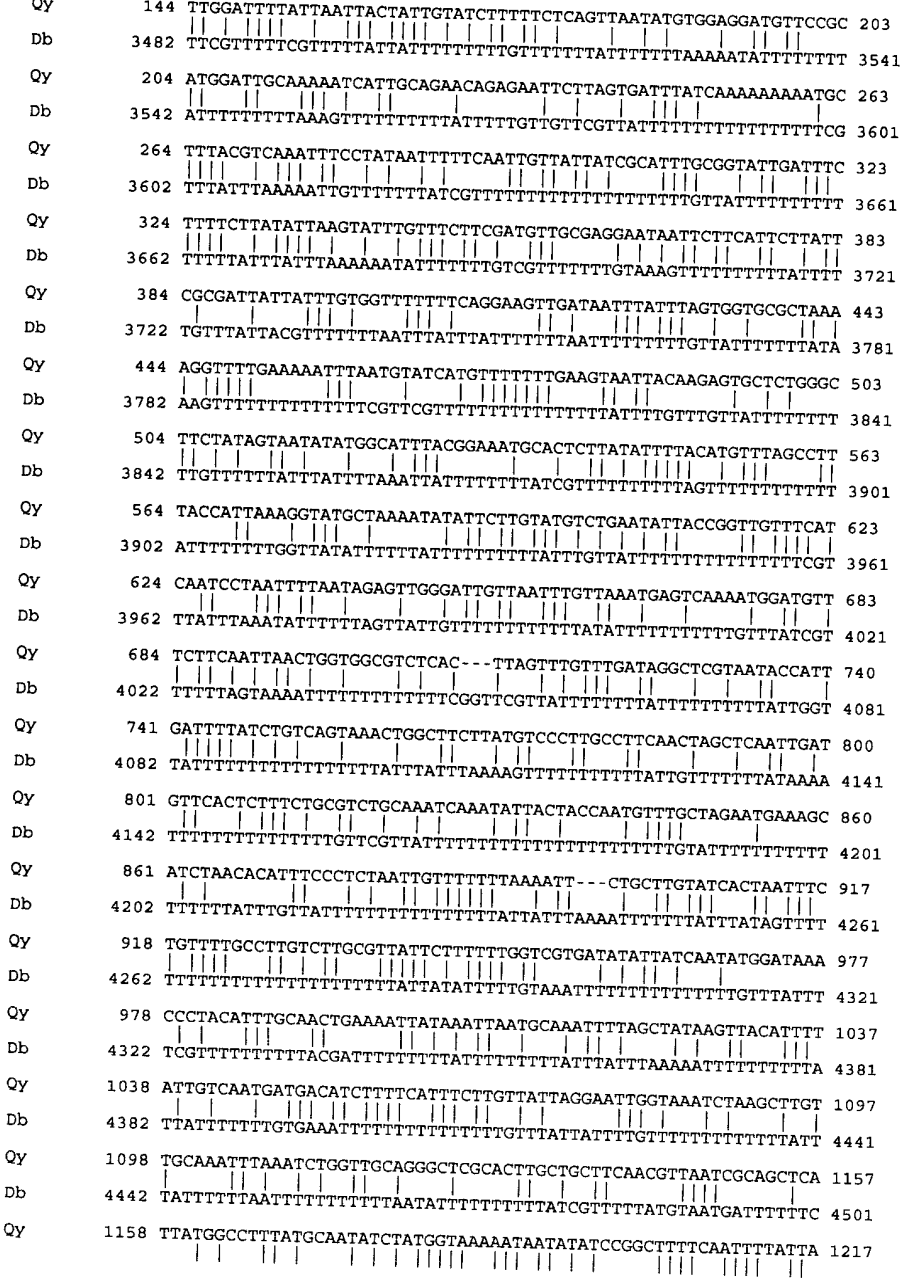
Qy       1104 TTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCAACGTTAATCGCAGCTCATTATGG 1163
              |||   | |  ||                |  |        ||        |  ||
Db      49661 TTTTNTTTTTTTTTTTTTTTNNNNNNNNNNTTTTNNTTTTTTTTNNTTTTTTTTTNNNN 49720

Qy       1164 CCTTTATGCAATATCTATGGTAAAAATAATATATCCGGCTTTTCAATTTTATTACCTTT 1222
              |||  |    | |              |  | |     ||||     || ||   |||
Db      49721 NNTTTTTTTTTTTTNNNNNNNTNTNTTTTTTTNNNNNNNTTTTTTNNNTTTTTTTTTTT 49779
```

```
RESULT 10
AX251493
LOCUS       AX251493                6292 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 461 from Patent WO0168912.
ACCESSION   AX251493
VERSION     AX251493.1  GI:15984916
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with tumor suppressor genes and
            oncogenes
  JOURNAL   Patent: WO 0168912-A 461 20-SEP-2001;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6292
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             7.3%;  Score 92.4;  DB 6;  Length 6292;
```

Best Local Similarity 43.9%; Pred. No. 0.00025;
Matches 494; Conservative 0; Mismatches 626; Indels 6; Gaps 2;



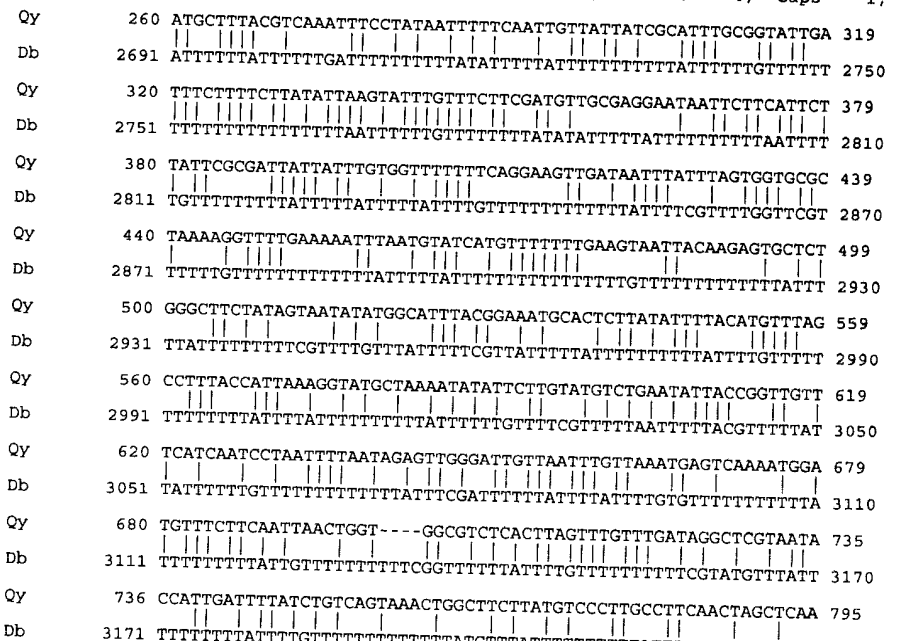
Db 4502 GGTTCGTTATTTTTTTTTTTTTTTGGTATTAATTATTTTTTTTATTTTTTTATTTATTTT 4561

Qy 1218 CCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTCTAT 1263

Db 4562 AAAATTATTTTTTTTTTTTTATTTTTTAGTTATATTATAGTGTTT 4607

RESULT 11
AX598853
LOCUS AX598853 3683 bp DNA linear PAT 14-FEB-2003
DEFINITION Sequence 193 from Patent WO02077272.
ACCESSION AX598853
VERSION AX598853.1 GI:28398991
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Berlin,K., Braun,A., Distler,J., Guetig,D., Howe,A., Mueller,J., Olek,A., Piepenbrock,C., Adorjan,P., Grabs,G., Lesche,R., Leu,E., Lewin,A., Lipscher,E., Maier,S., Model,F., Mueller,V., Otto,T., Pelet,C. and Ziebarth,H.
TITLE Methods and nucleic acids for the analysis of hematopoietic cell proliferative disorders
JOURNAL Patent: WO 02077272-A 193 03-OCT-2002;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
source 1. .3683
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Query Match 7.3%; Score 91.8; DB 6; Length 3683;
Best Local Similarity 45.3%; Pred. No. 0.00039;
Matches 374; Conservative 0; Mismatches 447; Indels 4; Gaps 1;



```
Qy        796 TTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATG 855
              || ||||| | | ||| |   | |     |    | ||  |    || |||  |  | ||
Db       3231 TTTATGTTTATTTTTTTTTTATTTTGTTTTTTTTTTTTTTATGTTTATTTTTTTTTATTT 3290

Qy        856 AAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATT 915
                | |       ||     |  || |||  ||    || |   | | |   |  ||
Db       3291 TGTTTTTTTTTTTTTATTTTAATTTTTTTTTCGTTTTTTTTTTTTTATTTTTATTTTTTT 3350

Qy        916 TCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATA 975
              | | ||||   || | ||   || || ||||| | |  | | | || | | | |   |
Db       3351 TTTTTTTTTATTTTTATTTGTTTTTTATTTTTAGTTTTTTTTGTTTTTTAATTTTTTTTT 3410

Qy        976 AACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATT 1035
                |  |||    ||   || |   ||  |    ||||  | |   ||   ||   |
Db       3411 TTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTGTTTTTTTTTTTATTGTTTTTTTTTA 3470

Qy       1036 TTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 1080
              || || |   |  |   |  |||| |      ||| ||    |||
Db       3471 TTGTTTTTTTTTTTTTTAGTTTTTGAAAATAAGTTTTTTTCGATT 3515
```

```
RESULT 12
CR936360/c
LOCUS       CR936360              154625 bp    DNA     linear   HTG 09-AUG-2005
DEFINITION  Homo sapiens chromosome X clone RP13-440N7, WORKING DRAFT SEQUENCE,
            2 unordered pieces.
ACCESSION   CR936360
VERSION     CR936360.11  GI:72060596
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 154625)
  AUTHORS   Whitehead,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 9, 2005 this sequence version replaced gi:71470739.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            -------------- Project Information
            Center project name: bB440N7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 154439 bases at least Q40
            Consensus quality: 154491 bases at least Q30
            Consensus quality: 154512 bases at least Q20
            Insert size: 154525; sum-of-contigs
            Insert size: 179913; 5.3% error; agarose-fp
            Quality coverage: 8.13x in Q20 bases; sum-of-contigs Quality
            coverage: 7.20x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   102638: contig of 102638 bp in length
            *  102639   102738: gap of 100 bp
            *  102739   154625: contig of 51887 bp in length.
FEATURES            Location/Qualifiers
     source         1. .154625
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="X"
                    /clone="RP13-440N7"
                    /clone_lib="RPCI-13.2"
     misc_feature   1. .102638
                    /note="assembly_fragment:00149
                    fragment_chain:1"
     misc_feature   102739. .154625
                    /note="assembly_fragment:02943
                    fragment_chain:1"
ORIGIN
```

```
  Query Match             7.2%;  Score 91;  DB 14;  Length 154625;
  Best Local Similarity   45.0%;  Pred. No. 0.00011;
  Matches  422;  Conservative    0;  Mismatches  510;  Indels    5;  Gaps    2;

Qy        144 TTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTGGAGGATGTTCCGC 203
              ||  |||||||| |||  |  | |    |||| |   || |||| |      || ||
Db      26908 TTATATTTTATTTATTTTTTATTTTATATTTTATTTTTTTATATTTTATTTATTTTATTA 26849

Qy        204 ATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGATTTATCAAAAAAAAATGC 263
                |  |||    | | |   |      |        |  ||  |  |||||      |
Db      26848 TTTTATTTTTATATTTTATTTTATATTTTATTTATTTTTTATTTTATATTTTATTTTATT 26789

Qy        264 TTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTC 323
              ||||  |  ||||  | | || ||| | || ||  | |   ||||    | ||  |||
Db      26788 TTTATATTTTATTTTATTTTATATTTTATTTATTTTATTTTTTATTTCATATTTTATTTTA 26729

Qy        324 TTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATT 383
              ||||  ||| || | | | ||  ||   | ||       | |  | | | |  |  ||  |
Db      26728 TTTTTATATTTTTATTTTATTTTATTTTATTTTTTATTTTACTTTTACATTTTATTTTAT 26669

Qy        384 CGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAA 443
                | ||||  | |  ||| |||     |  ||      ||  ||   |  |     | |
Db      26668 TTTATATTTTATATTTTATTTATTTATTTATTTTATATTTTATTTATTTTATTATTTTAT 26609

Qy        444 AGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGC 503
                |||  | | |||| | |||  | | |||  |    || ||    |   |  | | |
Db      26608 TTTATTTTTATATTTTATTTTATATTTTATTTATTTTTTATTTTATATTTTATTTTATTT 26549

Qy        504 TTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTT 563
              || |||  || | |||   || |                |  | |  | |     |||  ||
Db      26548 TTATATTTTATTTTATTTTATATTTTATTTATTTTTTATTTCATATTTTACTTTATTTTT 26489

Qy        564 TACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCAT 623
                || |   |||  ||   ||| || |  | |   |  | |||      || ||| ||
Db      26488 ATATTTTATTTTATTTTATTTTATTTTTTATTTTACTTTTACATTTTATTTTATTTTAT 26429

Qy        624 CAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTT 683
                |  || |||   ||    |   ||  || ||| ||  ||   | |       | ||
Db      26428 ATTTTATATTTTATTTATTTATTTATTTTATATTTTATTTTATATTTTATTTATTTATT 26369

Qy        684 TCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGAT 743
                || | || | |  |   | |  ||| || | |       |  | |  ||   || ||
Db      26368 ATTTTATTTTATTTTTATATTTTATTTTATATTTTATTTATTTTTTATTTTATATTTTAT 26309

Qy        744 TTTATCTGTCAGTAAACT-GGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATGT 802
              ||||| |   | | | |   || |||| ||   |     |     |   | |||  | |
Db      26308 TTTATTTTATATTTTATTCATTTTTTTATTTCATATTTTATTTTATTTTTATATTTTTAT 26249

Qy        803 TCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCAT 862
              |   | ||  |   | |   | | | | |||     || ||   |   ||   |   |
Db      26248 TTTATTTTATTTTATTTTTTATTTTACTTTTACATTTTATTTTATTTTATATTTTATATT 26189

Qy        863 CTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTTT 922
              ||    ||   | || ||  ||| ||  | || |  ||  ||   | ||||   |||
Db      26188 TTATTTATTTATTTTATATTTTATTTATTTTATTTTTTATTTCATATTTCATTTTATTTT 26129

Qy        923 TGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATAAACCCTA 982
              |   |  |  |   ||||| | ||||      |  ||||||  |   |      | |    |
```

```
Db      26128 TATATATTTATTTTTTATTTTATTTT----ATTTTATATTTTATTTTATTTTTATATTTT 26073

Qy        983 CATTTGCAACTGAAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTTTATTGT 1042
              ||||   | |   |  |||||  ||| |     ||||||  |       |||  | | |  |
Db      26072 TATTTTTTATTTTATTTTATATTTTATTTTTTATTTTATTTTATTTTTATTGTGTTTGTT 26013

Qy       1043 CAATGATGACATCTTTTCATTTCTTGTTATTAGGAAT 1079
              | |||  || || |  | | || | |  |  ||
Db      26012 ATGTTATGTTATGTTATGTTATGTTATGTTATGTTAT 25976


RESULT 13
AC138074
LOCUS       AC138074               66993 bp    DNA     linear   HTG 12-DEC-2002
DEFINITION  Homo sapiens chromosome 18 clone RP13-907D8 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC138074
VERSION     AC138074.1  GI:26553381
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 66993)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP13-907D8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 66993)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L28874
                Center clone name: 907_D_8
            --------------.
            * NOTE: This record contains 55 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.

*        1    1116: contig of 1116 bp in length
*     1117    1216: gap of 100 bp
*     1217    2336: contig of 1120 bp in length
*     2337    2436: gap of 100 bp
*     2437    3561: contig of 1125 bp in length
*     3562    3661: gap of 100 bp
*     3662    4781: contig of 1120 bp in length
*     4782    4881: gap of 100 bp
*     4882    6001: contig of 1120 bp in length
*     6002    6101: gap of 100 bp
*     6102    7244: contig of 1143 bp in length
*     7245    7344: gap of 100 bp
*     7345    8458: contig of 1114 bp in length
*     8459    8558: gap of 100 bp
*     8559    9640: contig of 1082 bp in length
*     9641    9740: gap of 100 bp
*     9741   10846: contig of 1106 bp in length
*    10847   10946: gap of 100 bp
*    10947   12103: contig of 1157 bp in length
*    12104   12203: gap of 100 bp
*    12204   13320: contig of 1117 bp in length
*    13321   13420: gap of 100 bp
*    13421   14520: contig of 1100 bp in length
*    14521   14620: gap of 100 bp
*    14621   15773: contig of 1153 bp in length
*    15774   15873: gap of 100 bp
*    15874   16991: contig of 1118 bp in length
*    16992   17091: gap of 100 bp
*    17092   18240: contig of 1149 bp in length
*    18241   18340: gap of 100 bp
*    18341   19476: contig of 1136 bp in length
*    19477   19576: gap of 100 bp
*    19577   20721: contig of 1145 bp in length
*    20722   20821: gap of 100 bp
*    20822   21942: contig of 1121 bp in length
*    21943   22042: gap of 100 bp
*    22043   23141: contig of 1099 bp in length
*    23142   23241: gap of 100 bp
*    23242   24372: contig of 1131 bp in length
*    24373   24472: gap of 100 bp
*    24473   25627: contig of 1155 bp in length
*    25628   25727: gap of 100 bp
*    25728   26888: contig of 1161 bp in length
*    26889   26988: gap of 100 bp
*    26989   28030: contig of 1042 bp in length
*    28031   28130: gap of 100 bp
*    28131   29303: contig of 1173 bp in length
*    29304   29403: gap of 100 bp
*    29404   30554: contig of 1151 bp in length
*    30555   30654: gap of 100 bp
*    30655   31751: contig of 1097 bp in length
*    31752   31851: gap of 100 bp
*    31852   32966: contig of 1115 bp in length
*    32967   33066: gap of 100 bp
*    33067   34113: contig of 1047 bp in length
*    34114   34213: gap of 100 bp
*    34214   35329: contig of 1116 bp in length
*    35330   35429: gap of 100 bp
*    35430   36574: contig of 1145 bp in length
*    36575   36674: gap of 100 bp
*    36675   37828: contig of 1154 bp in length
*    37829   37928: gap of 100 bp
*    37929   39063: contig of 1135 bp in length
*    39064   39163: gap of 100 bp
*    39164   40289: contig of 1126 bp in length
*    40290   40389: gap of 100 bp
*    40390   41471: contig of 1082 bp in length
*    41472   41571: gap of 100 bp
*    41572   42634: contig of 1063 bp in length
*    42635   42734: gap of 100 bp
*    42735   43904: contig of 1170 bp in length
*    43905   44004: gap of 100 bp
*    44005   45128: contig of 1124 bp in length
```

```
          *    45129    45228: gap of 100 bp
          *    45229    46358: contig of 1130 bp in length
          *    46359    46458: gap of 100 bp
          *    46459    47595: contig of 1137 bp in length
          *    47596    47695: gap of 100 bp
          *    47696    48785: contig of 1090 bp in length
          *    48786    48885: gap of 100 bp
          *    48886    49960: contig of 1075 bp in length
          *    49961    50060: gap of 100 bp
          *    50061    51132: contig of 1072 bp in length
          *    51133    51232: gap of 100 bp
          *    51233    52356: contig of 1124 bp in length
          *    52357    52456: gap of 100 bp
          *    52457    53556: contig of 1100 bp in length
          *    53557    53656: gap of 100 bp
          *    53657    54782: contig of 1126 bp in length
          *    54783    54882: gap of 100 bp
          *    54883    56008: contig of 1126 bp in length
          *    56009    56108: gap of 100 bp
          *    56109    57307: contig of 1199 bp in length
          *    57308    57407: gap of 100 bp
          *    57408    58529: contig of 1122 bp in length
          *    58530    58629: gap of 100 bp
          *    58630    59799: contig of 1170 bp in length
          *    59800    59899: gap of 100 bp
          *    59900    60983: contig of 1084 bp in length
          *    60984    61083: gap of 100 bp
          *    61084    62193: contig of 1110 bp in length
          *    62194    62293: gap of 100 bp
          *    62294    63397: contig of 1104 bp in length
          *    63398    63497: gap of 100 bp
          *    63498    64574: contig of 1077 bp in length
          *    64575    64674: gap of 100 bp
          *    64675    65800: contig of 1126 bp in length
          *    65801    65900: gap of 100 bp
          *    65901    66993: contig of 1093 bp in length.
FEATURES             Location/Qualifiers
     source          1. .66993
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP13-907D8"
                     /clone_lib="RPCI-13 Human Female BAC"
     gap             1117. .1216
                     /estimated_length=100
     gap             2337. .2436
                     /estimated_length=100
     gap             3562. .3661
                     /estimated_length=100
     gap             4782. .4881
                     /estimated_length=100
     gap             6002. .6101
                     /estimated_length=100
     gap             7245. .7344
                     /estimated_length=100
     gap             8459. .8558
                     /estimated_length=100
     gap             9641. .9740
                     /estimated_length=100
     gap             10847. .10946
                     /estimated_length=100
     gap             12104. .12203
                     /estimated_length=100
     gap             13321. .13420
                     /estimated_length=100

  Query Match             7.2%;  Score 90.6;  DB 14;  Length 66993;
  Best Local Similarity   38.5%;  Pred. No. 0.00017;
  Matches  324;  Conservative    0;  Mismatches  518;  Indels    0;  Gaps    0;

Qy        275 TTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATAT 334
```

```
Db      56368 TTNNTTTTNTTTTTTTTTNTTTTTTTTTTTTTNNNNTNTTTNNTNNTTTTNNNTTTTNNTTT 56427

Qy        335 TAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTA 394

Db      56428 TTNNTTTNTTTTTTTTTTTTTTTTNNCCCTTTTTTTTTNTTTTTTTTTNTNTTTTTTTN 56487

Qy        395 TTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAA 454

Db      56488 TTTTTTTTTTTTNTTNTTTTTTTTTTTTTTTTTNNNNTNNNNTNNTTTTTTTTTTTTTTT 56547

Qy        455 AATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAA 514

Db      56548 NNNNTNNTTTNTNTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTT 56607

Qy        515 TATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAG 574

Db      56608 TTTTTNTTTNTTTTTTTTTTNTTTTTTTTNTTTTTTTTTTTTTTTTTTTTNNTTTNTNTT 56667

Qy        575 GTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCATCAATCCTAATT 634

Db      56668 NTNTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTNTTTTTTTNTTTTTTTT 56727

Qy        635 TTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTTTCTTCAATTAA 694

Db      56728 TTTTTTNTTTTTTTTTTTTTNTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNT 56787

Qy        695 CTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGATTTTATCTGTCA 754

Db      56788 TTTTNTTTTTTTNNTTNTTTTTTTTTTTTNTTNTNTTTTTTTTTTTTTTTNNTNNTTNNT 56847

Qy        755 GTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATGTTCACTCTTTCTG 814

Db      56848 TTNTTTTTNTTNTTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTTNTTNNTNN 56907

Qy        815 CGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCATCTAACACATTTC 874

Db      56908 TNTTTTTTTNTTTTTTNTTTTNTTNTTTNTTTTTTTTTTTTTTTTNNTTTTNTTTTTTTTN 56967

Qy        875 CCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTTTTGCCTTGTCTTG 934

Db      56968 TTTTTTTTTTNNNTTTTTTTTNTTTTTTTTTTTTTTTTTNNNTTTTTTTTTTTTTTTTTTT 57027

Qy        935 CGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATAAACCCTACATTTGCAACTG 994

Db      57028 TTTTNNNTTTTTTTNTTTTTTTTTTTTTTTTTNTNNTNNNNNTTNTTTNNTTTTTTTTTTT 57087

Qy        995 AAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTTTATTGTCAATGATGACAT 1054

Db      57088 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTNTNTTTTTTTTTTTTTTTTNTTT 57147

Qy       1055 CTTTTCATTTCTTGTTATTAGGAATTGGTAAATCTAAGCTTGTTGCAAATTTAAATCTGG 1114

Db      57148 TTTTTTTTTTTNTTTTTTTNTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTT 57207

Qy       1115 TT 1116
              ||
Db      57208 TT 57209
```

```
RESULT 14
AC152242
LOCUS       AC152242              282289 bp    DNA     linear   HTG 01-JUL-2005
DEFINITION  Bos taurus clone CH240-7O10, *** SEQUENCING IN PROGRESS ***, 38
            unordered pieces.
ACCESSION   AC152242
VERSION     AC152242.3  GI:68227126
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
```

REFERENCE 1 (bases 1 to 282289)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 282289)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 282289)
AUTHORS Cow Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Jun 26, 2005 this sequence version replaced gi:58038019.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: FAGP
Center clone name: CH240-7O10
---------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 265527 bases at least Q40
Consensus quality: 268628 bases at least Q30
Consensus quality: 271489 bases at least Q20
Estimated insert size: 269724; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
---

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	18156: contig of 18156 bp in length
*	18157	18206: gap of 50 bp
*	18207	50971: contig of 32765 bp in length
*	50972	51021: gap of 50 bp
*	51022	62914: contig of 11893 bp in length
*	62915	62964: gap of 50 bp
*	62965	65254: contig of 2290 bp in length
*	65255	65304: gap of 50 bp
*	65305	66643: contig of 1339 bp in length
*	66644	66693: gap of 50 bp
*	66694	72399: contig of 5706 bp in length
*	72400	72821: gap of 422 bp
*	72822	74487: contig of 1666 bp in length
*	74488	74537: gap of 50 bp
*	74538	80785: contig of 6248 bp in length
*	80786	80835: gap of 50 bp
*	80836	87779: contig of 6944 bp in length
*	87780	87879: gap of unknown length
*	87880	92358: contig of 4479 bp in length
*	92359	92408: gap of 50 bp
*	92409	97196: contig of 4788 bp in length
*	97197	97246: gap of 50 bp
*	97247	104787: contig of 7541 bp in length
*	104788	104837: gap of 50 bp
*	104838	106100: contig of 1263 bp in length
*	106101	106200: gap of unknown length
*	106201	113418: contig of 7218 bp in length
*	113419	113606: gap of 188 bp
*	113607	115164: contig of 1558 bp in length
*	115165	115214: gap of 50 bp
*	115215	126597: contig of 11383 bp in length
*	126598	126697: gap of unknown length
*	126698	130816: contig of 4119 bp in length
*	130817	130866: gap of 50 bp
*	130867	133210: contig of 2344 bp in length
*	133211	134248: gap of 1038 bp
*	134249	137021: contig of 2773 bp in length
*	137022	138232: gap of 1211 bp
*	138233	139433: contig of 1201 bp in length
*	139434	139533: gap of unknown length
*	139534	144869: contig of 5336 bp in length
*	144870	144919: gap of 50 bp

```
            *   144920   151314: contig of 6395 bp in length
            *   151315   151364: gap of 50 bp
            *   151365   174159: contig of 22795 bp in length
            *   174160   174209: gap of 50 bp
            *   174210   216578: contig of 42369 bp in length
            *   216579   216628: gap of 50 bp
            *   216629   223479: contig of 6851 bp in length
            *   223480   223529: gap of 50 bp
            *   223530   229963: contig of 6434 bp in length
            *   229964   230118: gap of 155 bp
            *   230119   233635: contig of 3517 bp in length
            *   233636   233860: gap of 225 bp
            *   233861   245820: contig of 11960 bp in length
            *   245821   245870: gap of 50 bp
            *   245871   253648: contig of 7778 bp in length
            *   253649   253748: gap of unknown length
            *   253749   254859: contig of 1111 bp in length
            *   254860   254959: gap of unknown length
            *   254960   256040: contig of 1081 bp in length
            *   256041   256140: gap of unknown length
            *   256141   257392: contig of 1252 bp in length
            *   257393   257492: gap of unknown length
            *   257493   258889: contig of 1397 bp in length
            *   258890   258989: gap of unknown length
            *   258990   260305: contig of 1316 bp in length
            *   260306   260405: gap of unknown length
            *   260406   261931: contig of 1526 bp in length
            *   261932   262031: gap of unknown length
            *   262032   264065: contig of 2034 bp in length
            *   264066   264165: gap of unknown length
            *   264166   271493: contig of 7328 bp in length
            *   271494   271593: gap of unknown length
            *   271594   282289: contig of 10696 bp in length.
FEATURES             Location/Qualifiers
     source          1. .282289
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /clone="CH240-7O10"
     gap             18157. .18206
                     /estimated_length=50
     gap             50972. .51021
                     /estimated_length=50

  Query Match             7.2%;  Score 90.6;  DB 14;  Length 282289;
  Best Local Similarity   39.3%;  Pred. No. 9.7e-05;
  Matches  318;  Conservative    0;  Mismatches  492;  Indels    0;  Gaps    0;

Qy        264 TTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTC 323
              |||          |||     |  |||||    || || || |     |||     | ||  |||
Db      50119 TTTTTTNNNNTTTTNNNTTTTTTTTTTNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 50178

Qy        324 TTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATT 383
              ||||  || | ||    | |||  ||| ||       |         |  ||  ||  || || ||
Db      50179 TTTTTTTTTTTTTTTTTTTTTTTTTTTTNNNNTTTTTTTTTTTTTTTTTTTTTTTTTT 50238

Qy        384 CGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAA 443
                ||  | |   |  ||||||        ||  |  |||  |||  |  |    |
Db      50239 NNTTTTTNNTTTNNTTNNTTTTTTNNTTNNTTTTTTTTTTTTTTTTTTTTTTTTTTTT 50298

Qy        444 AGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGC 503
                ||||       |||  | | |   | |||||||        ||         |  | |
Db      50299 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTNNTTTTTTTTTTTTTTTTTTT 50358

Qy        504 TTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTT 563
              || | |  |  | | |    |||       |    | || | ||||  |  |||    ||
Db      50359 TTTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 50418

Qy        564 TACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCAT 623
              |    ||      |    |    | | || ||   | | |   | ||     || |||  |
Db      50419 TTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTNNNTTTTTTTTTTTTTTTTTTTTTTTTT 50478

Qy        624 CAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTT 683
                |  |    ||  |    ||    || ||  ||| ||   |   |     |   | ||
Db      50479 TTTTTTTTNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTT 50538

Qy        684 TCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGAT 743
              | ||   ||    | |   |  | |   ||   || |||  |     |  | |   ||  |
Db      50539 TTTTTTTTTTNNTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTNNTTTTTTTTTTTTTT 50598

Qy        744 TTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATGTT 803
              |||  | | |   |        ||    | |   ||   ||          |   ||  | ||
Db      50599 TTTTTTTTTTTTTNNTNNTTTTTTNNNTTTNNTTTTTTTNNNNNNNTTTTTTNNTTTT 50658

Qy        804 CACTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCATC 863
                | |||  |                  | ||  |      | |||  |
Db      50659 TTTTTTTTTTTTTNNNTTTNNNTTTTTTTTTTNNTTTTTTTTTTTTTTTNNTTTTNNNN 50718

Qy        864 TAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTTTT 923
              |      |||||||    || |||  ||    || |  || | |      ||| | ||||
Db      50719 TTTTTTTTTCCCTTNNNTTTTTNNTTTTTTTTTTNNTTTTTTNNNNTTTTTTTTTTTT 50778

Qy        924 GCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATAAACCCTAC 983
              || |      |  || ||||||  |  |  | | || |   | |  |      |
Db      50779 TTTTTTTTNNNNTTNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTTTTTT 50838

Qy        984 ATTTGCAACTGAAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTTTATTGTC 1043
                |||     |        |   ||        ||||   |      ||   |  | || |
Db      50839 TTTTTTTTTTTTTTTNNTTNNTTNNTTTTTTTTTTTTTTTNNTTTTTTTNNTTTTTT 50898

Qy       1044 AATGATGACATCTTTTCATTTCTTGTTATT 1073
                  |   | ||||  ||| ||   ||
Db      50899 TTNNNTTTTTTTTTTTTTTTTTTNNNTTT 50928


RESULT 15
AX346599
LOCUS       AX346599                6668 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 1670 from Patent WO0200928.
ACCESSION   AX346599
VERSION     AX346599.1  GI:18494485
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 1670 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6668
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

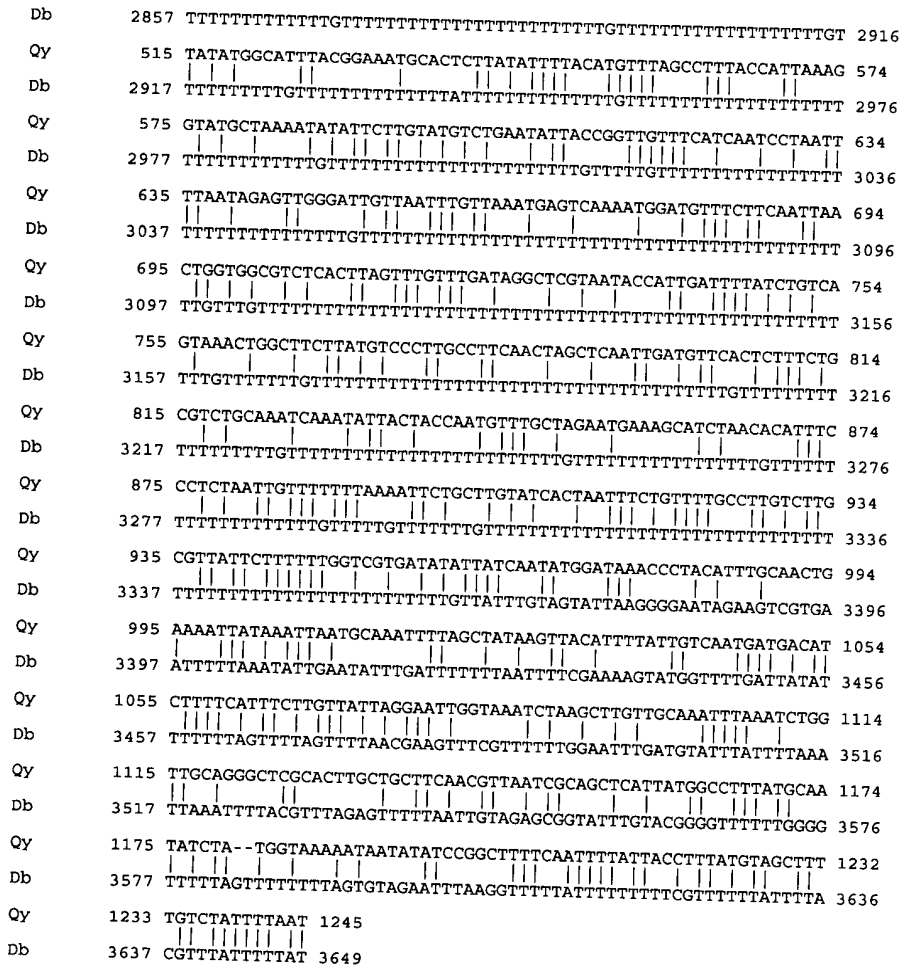
  Query Match             7.1%;  Score 90.2;  DB 6;  Length 6668;
  Best Local Similarity   44.0%;  Pred. No. 0.00051;
  Matches  428;  Conservative    0;  Mismatches  543;  Indels    2;  Gaps    1;

Qy        275 TTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATAT 334
              |||  | |  |||||   || || || |   ||||    | ||  ||| |||| || | |
Db       2677 TTTTTTTTTTTTTTTTTGTTTTTTTTTTTTGATTTTTTTTTTTTTTTTTGTTTTTTTTTT 2736

Qy        335 TAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTA 394
              |  || ||| ||| ||    | ||          |  |  ||  || || ||     || ||
Db       2737 TTTGTGTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTTTT 2796

Qy        395 TTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAA 454
              ||  |  |||||||        ||  |  |||  |||  | ||    |    ||||
Db       2797 TTGTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTCGTTTTTTTTTTTTTTTTTGT 2856

Qy        455 AATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAA 514
                |||  | | |  | |||||||    |  ||      ||  | |    || | |  |
```



Search completed: May 10, 2006, 20:34:17
Job time : 4602.32 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:42:47 ; Search time 590.987 Seconds
(without alignments)
14276.972 Million cell updates/sec

Title: US-09-423-093B-1_COPY_8646_9911
Perfect score: 1266
Sequence: 1 atggtattaacagtgaaaaa.....gagcgaaaaatgtctattga 1266

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1266	100.0	14516	2	AAX06748	Aax06748 E. coli O
	2	1266	100.0	14516	3	AAZ56375	Aaz56375 Escherich
	3	96	7.6	3683	8	ABZ10199	Abz10199 Haematopo
	4	92.4	7.3	6292	4	AAS46735	Aas46735 Tumour su
	5	91.8	7.3	3683	8	ABZ10053	Abz10053 Haematopo
	6	90.2	7.1	6668	6	ABL33697	Abl33697 Human imm
c	7	86.4	6.8	1000	12	ADQ62833	Adq62833 Homopoly-
	8	86.4	6.8	1000	12	ADQ62832	Adq62832 Homopoly-
c	9	86.4	6.8	1300	12	ADP85917	Adp85917 Synthetic
	10	82.2	6.5	7442	4	AAS46686	Aas46686 Tumour su
	11	80.4	6.4	6314	6	ABL54314	Abl54314 Chemicall
	12	78	6.2	14006	6	ABL33958	Abl33958 Human imm
c	13	77.4	6.1	1039	14	ACL64817	Acl64817 M. xanthu
c	14	76.6	6.1	6644	2	AAX33181	Aax33181 Base sequ
c	15	76.6	6.1	7372	2	AAX33182	Aax33182 Base sequ
c	16	76.6	6.1	7797	2	AAX33180	Aax33180 Cowpox vi
c	17	76.6	6.1	7996	2	AAX33184	Aax33184 Base sequ
	18	76.4	6.0	5963	6	ABL34613	Abl34613 Human met
	19	76.4	6.0	5963	7	ADS99874	Ads99874 Complemen
	20	76.2	6.0	15479	6	ABK39964	Abk39964 Human che
	21	75.4	6.0	5750	4	AAS46708	Aas46708 Tumour su
	22	75.4	6.0	5750	6	ABL34008	Abl34008 Human imm
	23	74.2	5.9	6113	6	ABL33823	Abl33823 Human imm
	24	74.2	5.9	15387	6	ABL32184	Abl32184 Human imm
	25	74	5.8	5145	6	ABL32348	Abl32348 Human imm
	26	74	5.8	5145	6	ABL34464	Abl34464 Human met
	27	74	5.8	5145	7	ADS99725	Ads99725 Bisulphit
	28	73.4	5.8	11555	6	ABL32616	Abl32616 Human imm
	29	73.4	5.8	11555	6	AAD28380	Aad28380 Human che
	30	73.2	5.8	5893	6	ABL32859	Abl32859 Human imm
	31	73	5.8	6767	4	AAS46608	Aas46608 Tumour su
	32	73	5.8	17131	6	ABL33053	Abl33053 Human imm
	33	72.8	5.8	5493	13	ADS89675	Ads89675 Oligonucl
	34	72.2	5.7	16766	6	ABL34157	Abl34157 Human imm
	35	71.4	5.6	6898	6	ABN80222	Abn80222 Human che
	36	71.4	5.6	8666	10	ADB54239	Adb54239 Pretreate
	37	71.4	5.6	8666	10	ADE84177	Ade84177 Human lym
	38	71.4	5.6	8666	13	ADS89541	Ads89541 Oligonucl
	39	71.4	5.6	9539	4	AAS45347	Aas45347 Chemicall
	40	71.4	5.6	9539	6	ABK28180	Abk28180 DNA trans
c	41	71.2	5.6	612	14	ACL64894	Acl64894 M. xanthu
	42	71.2	5.6	6291	4	AAS46451	Aas46451 Tumour su
	43	71	5.6	11812	4	AAS45502	Aas45502 Chemicall
	44	71	5.6	11812	4	AAS46742	Aas46742 Tumour su
	45	71	5.6	11812	6	ABL34119	Abl34119 Human imm

ALIGNMENTS

RESULT 1
AAX06748
ID AAX06748 standard; DNA; 14516 BP.
XX
AC AAX06748;
XX
DT 26-APR-1999 (first entry)
XX
DE E. coli O111 antigen gene cluster.
XX
KW O antigen; O111 antigen; diagnosis; wbdH gene; wzx gene; wzy gene;
KW wbdM gene; rfb gene; transferase; polymerase; flippase; diarrhoea;
KW haemorrhagic colitis; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 739. .1932
FT /*tag= a
FT /product= "galactosyl_transferase"
FT /note= "wbdH gene (ORF1), this region is specifically
FT claimed in Claim 6, and encodes AAW88300"
FT CDS 1983. .3101
FT /*tag= b
FT /note= "gmd gene (ORF2); encodes AAW88301"
FT CDS 3174. .3623
FT /*tag= c
FT /note= "wbdI gene (ORF3); encodes AAW88302"
FT CDS 3657. .5105
FT /*tag= d
FT /note= "manC gene (ORF4); encodes AAW88303"
FT CDS 5114. .6484
FT /*tag= e
FT /note= "manB gene (ORF5); encodes AAW88304"
FT CDS 6521. .7444
FT /*tag= f
FT /note= "wbdJ gene (ORF6); encodes AAW88305"
FT CDS 7441. .8607
FT /*tag= g
FT /note= "wbdK gene (ORF7); encodes AAW88306"
FT CDS 8646. .9911
FT /*tag= h

```
FT                   /product= "O_antigen_flippase"
FT                   /note= "wzx gene (ORF8), this region is specifically
FT                   claimed in Claim 6; encodes AAW88307"
FT   CDS             9901. .10953
FT                   /*tag=  i
FT                   /product= "O_antigen_polymerase"
FT                   /note= "wzy gene (ORF9), this region is specifically
FT                   claimed in Claim 6; encodes AAW88308"
FT   CDS             10931. .11824
FT                   /*tag=  j
FT                   /product= "colitose_or_glucose_transferase"
FT                   /note= "wbdL gene (ORF10); encodes AAW88309"
FT   CDS             11821. .12945
FT                   /*tag=  k
FT                   /product= "colitose or glucose transferase"
FT                   /note= "wbdM gene (ORF11), this region is specifically
FT                   claimed in Claim 6; encodes AAW88310"
FT   CDS             13124. .14516
FT                   /*tag=  l
FT                   /note= "gnd gene (partial sequence); encodes AAW88311"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;    98WO-AU000315.
XX
PR   01-MAY-1997;    97AU-00006545.
PR   22-JUL-1997;    97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88300, AAW88301, AAW88302, AAW88303, AAW88304, AAW88305,
DR   AAW88306, AAW88307, AAW88308, AAW88309, AAW88310, AAW88311.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 6; Fig 7; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Escherichia coli O111 O antigen. The gene cluster was
CC   isolated from a cosmid gene bank of E. coli M92 chromosomal DNA by
CC   screening with E. coli O111 antiserum. Open reading frames within the
CC   gene cluster were identified on the basis of homology to known sequences.
CC   The genes encode proteins (see AAW88300-11) involved the synthesis of
CC   sugars present in the polysaccharide antigen, and in the transport or
CC   processing of polysaccharide or oligosaccharide units. The use of nucleic
CC   acid molecules derived from particular assembly and transport genes,
CC   particularly wbd (transferase), wzx (flippase) and wzy (polymerase)
CC   genes, within O antigen gene clusters improves the specificity of methods
CC   for the detection and identification of O antigens, e.g. in testing food-
CC   or faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of enteropathogenic E. coli strains that cause
CC   diarrhoea and haemorrhagic colitis
XX
SQ   Sequence 14516 BP; 4582 A; 2221 C; 2971 G; 4742 T; 0 U; 0 Other;
```

```
  Query Match           100.0%;  Score 1266;  DB 2;  Length 14516;
  Best Local Similarity  100.0%;  Pred. No. 1e-234;
  Matches 1266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8646 ATGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTT 8705

Qy         61 ATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8706 ATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCAC 8765

Qy        121 CTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8766 CTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAG 8825

Qy        181 TTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8826 TTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTT 8885

Qy        241 AGTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8886 AGTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATT 8945

Qy        301 ATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8946 ATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCG 9005

Qy        361 AGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9006 AGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGAT 9065

Qy        421 AATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9066 AATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAA 9125

Qy        481 GTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9126 GTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTC 9185

Qy        541 TTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9186 TTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGT 9245

Qy        601 CTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9246 CTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTG 9305

Qy        661 TTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9306 TTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTT 9365

Qy        721 GATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9366 GATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGC 9425

Qy        781 CTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9426 CTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCA 9485

Qy        841 ATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9486 ATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTG 9545

Qy        901 CTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9546 CTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATA 9605

Qy        961 TTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9606 TTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTA 9665

Qy       1021 GCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9666 GCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 9725

Qy       1081 GGTAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9726 GGTAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCA 9785

Qy       1141 ACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9786 ACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCG 9845
```

```
Qy       1201 GCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9846 GCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTC 9905

Qy       1261 TATTGA 1266
             ||||||
Db       9906 TATTGA 9911


RESULT 2
AAZ56375
ID   AAZ56375 standard; DNA; 14516 BP.
XX
AC   AAZ56375;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:45.
XX
KW   Flagellin; fliC; antigen; detection; ds.
XX
OS   Escherichia coli.
XX
PN   WO9961458-A1.
XX
PD   02-DEC-1999.
XX
PF   21-MAY-1999;   99WO-AU000385.
XX
PR   21-MAY-1998;   98AU-00003634.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 2000-072598/06.
XX
PT   Novel nucleic acid molecule useful for the detection of flagellated
PT   bacterial strains in food, feces, etc.
XX
PS   Claim 3; Page 216-221; 245pp; English.
XX
CC   AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all or
CC   part of an Escherichia coli flagellin protein except a protein expressed
CC   by E. coli H1, H7, H12 or H48 type strains. The present invention also
CC   describes a method of detecting the presence of E. coli of a particular H
CC   serotype in a sample, comprising specifically hybridising a nucleic acid,
CC   preferably at least a pair, derived from a flagellating gene, specific
CC   for a particular flagellin gene associated with the H serotype, to any
CC   E.coli in the sample which contain the gene, and detecting any hybridised
CC   molecules, identifying the presence of that serotype in the sample. (I)
CC   are useful for: (1) detecting the presence of E. coli of H serotype in a
CC   sample by hybridising at least one or a pair of (I) to any E. coli in the
CC   sample and detecting the hybridised nucleic acid molecules; and (2) for
CC   detecting the presence of both O and H-serotypes of E. coli by
CC   hybridising at least one or a pair of (I) to any E. coli present in the
CC   sample and detecting the hybridised nucleic acid molecules. (I) is
CC   particularly useful for detecting the combination of O and H antigen.
CC   Hybridised (I) when using at least one (I) is detected by southern blot
CC   analysis and, when using a pair of (I), is detected by polymerase chain
CC   reaction (PCR). AAZ56399 to AAZ56420 represent primers used in the
CC   exemplification of the present invention
XX
SQ   Sequence 14516 BP; 4582 A; 2221 C; 2971 G; 4742 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1266;  DB 3;  Length 14516;
  Best Local Similarity   100.0%;  Pred. No. 1e-234;
  Matches 1266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8646 ATGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTT 8705

Qy         61 ATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8706 ATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCAC 8765

Qy        121 CTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8766 CTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAG 8825

Qy        181 TTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8826 TTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTT 8885

Qy        241 AGTGATTTATCAAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8886 AGTGATTTATCAAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATT 8945

Qy        301 ATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8946 ATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCG 9005

Qy        361 AGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9006 AGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGAT 9065

Qy        421 AATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9066 AATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAA 9125

Qy        481 GTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9126 GTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTC 9185

Qy        541 TTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9186 TTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGT 9245

Qy        601 CTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9246 CTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTG 9305

Qy        661 TTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9306 TTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTT 9365

Qy        721 GATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9366 GATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGC 9425

Qy        781 CTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9426 CTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCA 9485

Qy        841 ATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9486 ATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTG 9545

Qy        901 CTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9546 CTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATA 9605

Qy        961 TTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9606 TTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTA 9665

Qy       1021 GCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9666 GCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 9725

Qy       1081 GGTAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9726 GGTAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCA 9785

Qy       1141 ACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCG 1200
```

```
Db       9786 ACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCG 9845

Qy       1201 GCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTC 1260

Db       9846 GCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTC 9905

Qy       1261 TATTGA 1266

Db       9906 TATTGA 9911
```

RESULT 3
ABZ10199
ID ABZ10199 standard; DNA; 3683 BP.
XX
AC ABZ10199;
XX
DT 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #339.
XX
KW Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 339; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients

XX
SQ Sequence 3683 BP; 716 A; 0 C; 760 G; 2207 T; 0 U; 0 Other;

```
  Query Match             7.6%;  Score 96;  DB 8;  Length 3683;
  Best Local Similarity   45.3%;  Pred. No. 4e-09;
  Matches  428;  Conservative    0;  Mismatches  510;  Indels    6;  Gaps    2;

Qy        141 TAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTGGAGGATGTTC 200

Db       2574 TAGTTAGTTTTTTGTTTTTTTATTTGTTTATTGTATTTTTTTTTTTTTATTTTGTTTTTAT 2633

Qy        201 CGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGATTTATCAAAAAAAAA 260

Db       2634 TTTATTGTTTGGTGGGAAAGGTTGTAGTTATTATGTATTTATTTATTTTTATTTTTTATT 2693

Qy        261 TGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGAT 320

Db       2694 TTTTATTTTTTGATTTTTTTTTTATATTTTTATTTTTTTTTT--TATTTTTTGTTTTTTT 2751

Qy        321 TTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTT 380

Db       2752 TTTTTTTTTTTTTTTTAATTTTTTGTTTTTTTTATATATTTTTATTTTTTTTTTAATTTTT 2811

Qy        381 ATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCT 440

Db       2812 GTTTTTTTTTATTTTTATTTTTATTTTGTTTTTTTTTTTTTATTTTTGTTTTGGTTTGTT 2871

Qy        441 AAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTG 500

Db       2872 TTTTGTTTTTTTTTTTTTATTTTTATTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTATTTT 2931

Qy        501 GGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGC 560

Db       2932 TATTTTTTTTTTGTTTTGTTTATTTTTTGTTATTTTTATTTTTTTTTTTATTTTGTTTTTT 2991

Qy        561 CTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTT 620

Db       2992 TTTTTTTATTTTATTTTTTTTTTATTTTTTGTTTTTGTTTTTAATTTTTATGTTTTTATT 3051

Qy        621 CATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGAT 680

Db       3052 ATTTTTTGTTTTTTTTTTTTTTATTTTGATTTTTTATTTTATTTTGTGTTTTTTTTTTTAT 3111

Qy        681 GTTTCTTCA----ATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATAC 736

Db       3112 TTTTTTTTATTGTTTTTTTTTTTGGTTTTTTATTTTGTTTTTTTTTTTGTATGTTTATTT 3171

Qy        737 CATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAAT 796

Db       3172 TTTTTTTATTTTGTTTTTTTTTTTTTTATGTTTATTTTTTTTTTATTTTGTTTTTTTTTTT 3231

Qy        797 TGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGA 856

Db       3232 TTATGTTTATTTTTTTTTTTATTTTGTTTTTTTTTTTTTTATGTTTATTTTTTTTTTATTTT 3291

Qy        857 AAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTT 916

Db       3292 GTTTTTTTTTTTTTATTTTTAATTTTTTTTTTGTTTTTTTTTTTTTATTTTTATTTTTTT 3351

Qy        917 CTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATAA 976

Db       3352 TTTTTTTTATTTTTATTTGTTTTTTATTTTTAGTTTTTTTTGTTTTTTAATTTTTTTTTT 3411

Qy        977 ACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTT 1036

Db       3412 TTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTGTTTTTTTTTTTATTGTTTTTTTTTAT 3471

Qy       1037 TATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 1080

Db       3472 TGTTTTTTTTTTTTTTAGTTTTTGAAAATAAGTTTTTTTTGATT 3515
```

RESULT 4

AAS46735
ID AAS46735 standard; DNA; 6292 BP.
XX
AC AAS46735;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #459.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 459; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6292 BP; 1736 A; 99 C; 1304 G; 3153 T; 0 U; 0 Other;

Query Match 7.3%; Score 92.4; DB 4; Length 6292;
Best Local Similarity 43.9%; Pred. No. 2.1e-08;
Matches 494; Conservative 0; Mismatches 626; Indels 6; Gaps 2;

```
Qy        144 TTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTGGAGGATGTTCCGC 203
              || | ||||  |  ||| |||| | | || || |    |  | |    |  || ||
Db       3482 TTCGTTTTTCGTTTTTATTATTTTTTTTTGTTTTTTTATTTTTTTAAAAATATTTTTTT 3541

Qy        204 ATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGATTTATCAAAAAAAAATGC 263
              ||   ||   ||| |  ||     |       | |   |  ||| |          |

Db       3542 ATTTTTTTTTAAAGTTTTTTTTTTATTTTTGTTGTTCGTTATTTTTTTTTTTTTTTTTCG 3601

Qy        264 TTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTC 323
              ||||  | ||| ||  | |  | |      || || || |    ||||   | ||  |||
Db       3602 TTTATTTAAAAATTGTTTTTTTATCGTTTTTTTTTTTTTTTTTTTTTGTTATTTTTTTTT 3661

Qy        324 TTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATT 383
              |||| | |||| |  | | ||| || |  |||      | | | || ||  ||  | ||
Db       3662 TTTTTATTTATTTAAAAAATATTTTTTTGTCGTTTTTTTGTAAAGTTTTTTTTTTATTTT 3721

Qy        384 CGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAA 443
              |   |   ||| |   ||| |      || |  ||| ||| |  |    || |
Db       3722 TGTTTATTACGTTTTTTTAATTTATTTATTTTTTTAATTTTTTTTTGTTATTTTTTTATA 3781

Qy        444 AGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGC 503
              | |||||      |||    |    | |||||||    || ||       |  | |
Db       3782 AAGTTTTTTTTTTTTTTCGTTCGTTTTTTTTTTTTTTTTATTTTGTTTGTTATTTTTTTT 3841

Qy        504 TTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTT 563
              || | |  || |    |  | |||       |  | || | |||||  | |||   ||
Db       3842 TTGTTTTTTATTTATTTTAAATTATTTTTTTTATCGTTTTTTTTTTTAGTTTTTTTTTTTT 3901

Qy        564 TACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCAT 623
                 ||   | |||  |      | || || ||| | | | | ||    || |||| |
Db       3902 ATTTTTTTTGGTTATATTTTTTATTTTTTTTTTTATTTGTTATTTTTTTTTTTTTTTCGT 3961

Qy        624 CAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTT 683
              ||   ||| ||  |    |  | ||  ||  |||  ||  |   |      |  ||  |
Db       3962 TTATTTAAATATTTTTTAGTTATTGTTTTTTTTTTTTATATTTTTTTTTTTGTTTATCGT 4021

Qy        684 TCTTCAATTAACTGGTGGCGTCTCAC---TTAGTTTGTTTGATAGGCTCGTAATACCATT 740
              | || | | || | |  |    | | |   |   | | |||  ||   |  | |  ||   |
Db       4022 TTTTTAGTAAAATTTTTTTTTTTTTCGGTTCGTTATTTTTTTTATTTTTTTTTTATTGGT 4081

Qy        741 GATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGAT 800
              |||||  | | |    |    |  ||  | ||   ||   ||    |   |  ||  |
Db       4082 TATTTTTTTTTTTTTTTTTTATTTATTTAAAAGTTTTTTTTTTTATTGTTTTTTTATAAAA 4141

Qy        801 GTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGC 860
              ||   | ||| |  ||  |  | |     | ||  |      | ||||      |
Db       4142 TTTTTTTTTTTTTTTGTTCGTTATTTTTTTTTTTTTTTTTTTTTTGTATTTTTTTTTTT 4201

Qy        861 ATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATT---CTGCTTGTATCACTAATTTC 917
              | |        ||   | |  || ||||||   | ||    |  || |||   || |||
Db       4202 TTTTTTATTTGTTATTTTTTTTTTTTTTTTTTATTATTTAAAATTTTTTTATTTATAGTTTT 4261

Qy        918 TGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATAAA 977
              | ||||  || | ||  ||||| | |||| ||    | | || |   |
Db       4262 TTTTTTTTTTTTTTTTTTTTTATTATATTTTTGTAAATTTTTTTTTTTTTTTGTTTATTT 4321

Qy        978 CCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTTT 1037
              | |   |||   ||       || | | ||   |  | ||  |   | |  ||   |||
Db       4322 TCGTTTTTTTTTTACGATTTTTTTTTATTTTTTTTTATTTATTTAAAAATTTTTTTTTTA 4381

Qy       1038 ATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATTGGTAAATCTAAGCTTGT 1097
              | |   |  ||| || |||| ||| ||  | |       ||| |   | |    | |
Db       4382 TTATTTTTTTGTGAAATTTTTTTTTTTTTTTTGTTTATTATTTTGTTTTTTTTTTTTATT 4441

Qy       1098 TGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCAACGTTAATCGCAGCTCA 1157
              |     || |  | |  ||   |     |      ||  |  ||      ||||      |
Db       4442 TATTTTTTAATTTTTTTTTTTTAATATTTTTTTTTTTATCGTTTTTATGTAATGATTTTTTC 4501

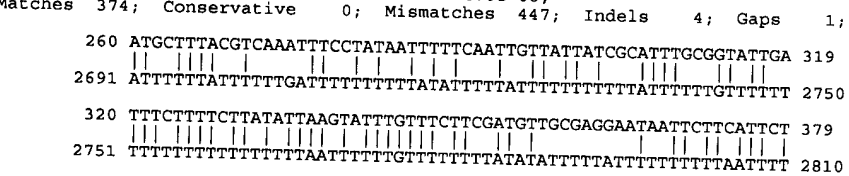
Qy       1158 TTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCGGCTTTTCAATTTTATTA 1217
              | |   || |    | | | |||||  ||| || | |      ||||  ||||  ||
Db       4502 GGTTCGTTATTTTTTTTTTTTTTTGGTATTAATTATTTTTTTTATTTTTTTATTTATTTT 4561

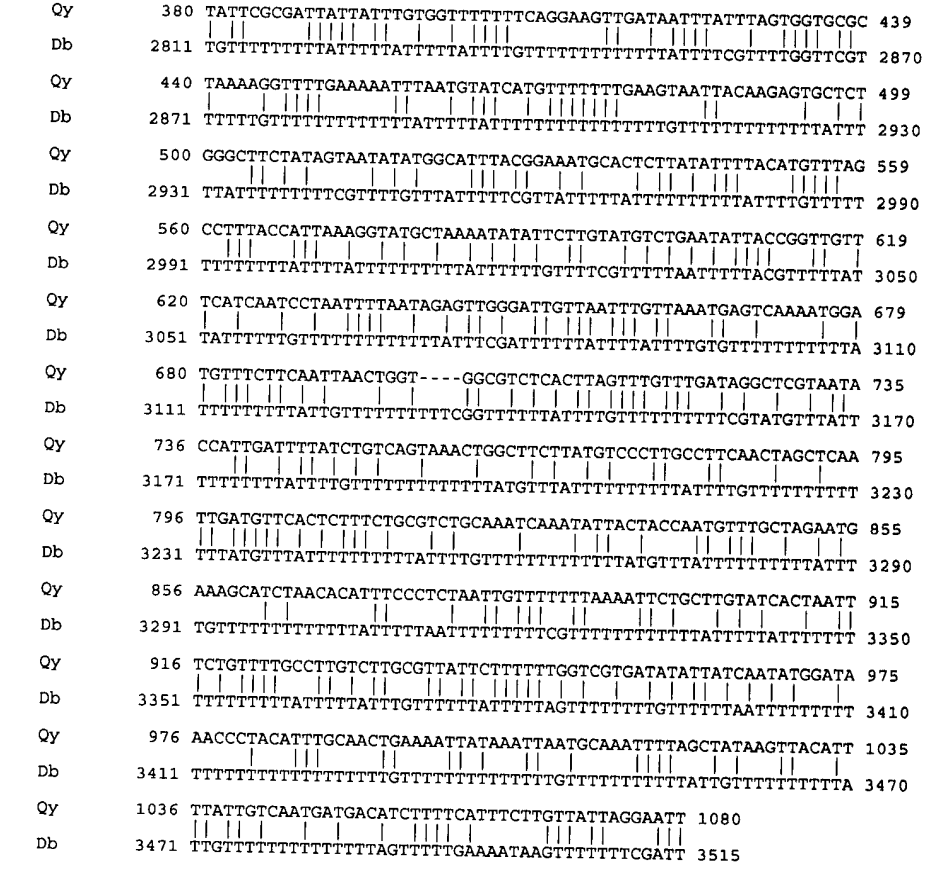
Qy       1218 CCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTCTAT 1263
                 |   |  |||| | ||||||  |||    |  |  || | |
Db       4562 AAAATTATTTTTTTTTTTTTTATTTTTTTAGTTATATTATAGTGTTT 4607
```

RESULT 5
ABZ10053

ID ABZ10053 standard; DNA; 3683 BP.
XX
AC ABZ10053;
XX
DT 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #193.
XX
KW Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 193; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 3683 BP; 716 A; 108 C; 760 G; 2099 T; 0 U; 0 Other;

Query Match 7.3%; Score 91.8; DB 8; Length 3683;
Best Local Similarity 45.3%; Pred. No. 2.6e-08;
Matches 374; Conservative 0; Mismatches 447; Indels 4; Gaps 1;





RESULT 6
ABL33697
ID ABL33697 standard; DNA; 6668 BP.
XX
AC ABL33697;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1670.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX

PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1670; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6668 BP; 1329 A; 328 C; 1736 G; 3274 T; 0 U; 1 Other;

Query Match 7.1%; Score 90.2; DB 6; Length 6668;
Best Local Similarity 44.0%; Pred. No. 5.5e-08;
Matches 428; Conservative 0; Mismatches 543; Indels 2; Gaps 1;

```
Qy        275 TTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATAT 334
              |||  | |  |||||   || || || |   ||||    | ||  ||| |||| || | |
Db       2677 TTTTTTTTTTTTTTTTTGTTTTTTTTTTTTGATTTTTTTTTTTTTTTTTGTTTTTTTTTT 2736

Qy        335 TAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTA 394
              |  || ||| ||| ||   | ||         |  |  ||  || || ||      || ||
Db       2737 TTTGTGTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTT 2796

Qy        395 TTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAA 454
              ||  |  |||||||        ||  |  ||| |||  | ||     |       ||||
Db       2797 TTGTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTCGTTTTTTTTTTTTTTTTTTGT 2856

Qy        455 AATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAA 514
                |||  | | |  | |||||||    |  ||      ||  | |    || | |  |
Db       2857 TTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTGT 2916

Qy        515 TATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAG 574
              | | |      ||        |       || | ||||   |||||    |||    ||
Db       2917 TTTTTTTTGTTTTTTTTTTTTTTATTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTT 2976

Qy        575 GTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCATCAATCCTAATT 634
              | | |    | | || || | | | |    | ||     ||||||  |   |  | ||
Db       2977 TTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTGTTTTGTTTTTTTTTTTTTTTT 3036

Qy        635 TTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTTTCTTCAATTAA 694
              ||  |    ||      | ||  ||| ||   |    |      |   | ||| ||   ||
Db       3037 TTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3096

Qy        695 CTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGATTTTATCTGTCA 754
              | | |  | |    ||  ||| |||  |     |  |  |     ||  |||| | | |
Db       3097 TTGTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3156

Qy        755 GTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATGTTCACTCTTTCTG 814
              |    |   |  || | |   ||   ||    |   |  ||  | ||   | ||| |
Db       3157 TTTGTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTT 3216

Qy        815 CGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCATCTAACACATTTC 874
              | |     |  | ||  |     | |||  |    |       |  |       |||
Db       3217 TTTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTGTTTTTT 3276

Qy        875 CCTCTAATTGTTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTTTTGCCTTGTCTTG 934
              | |  || ||| |||    || |   | | |   |  ||| | ||||   || | ||
Db       3277 TTTTTTTTTTTTTGTTTTGTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3336

Qy        935 CGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATAAACCCTACATTTGCAACTG 994
              || || ||||||  |  |  | | ||||   ||    |||     | |   |
Db       3337 TTTTTTTTTTTTTTTTTTTTTTTTGTTATTTGTAGTATTAAGGGGAATAGAAGTCGTGA 3396

Qy        995 AAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTTTATTGTCAATGATGACAT 1054
              |   ||| | ||| |         ||   |  |  ||  |       ||    |||| | ||
Db       3397 ATTTTTAAATATTGAATATTTGATTTTTTTAATTTTCGAAAAGTATGGTTTTGATTATAT 3456

Qy       1055 CTTTTCATTTCTTGTTATTAGGAATTGGTAAATCTAAGCTTGTTGCAAATTTAAATCTGG 1114
              ||||  | || | ||| | |  ||| |      | |  |  | |      |||||  |
Db       3457 TTTTTTAGTTTTAGTTTTAACGAAGTTTCGTTTTTTGGAATTTGATGTATTTATTTTAAA 3516

Qy       1115 TTGCAGGGCTCGCACTTGCTGCTTCAACGTTAATCGCAGCTCATTATGGCCTTTATGCAA 1174
              ||  |      ||       |  || |  ||  | ||     |  |   ||  ||| ||
Db       3517 TTAAATTTTACGTTTAGAGTTTTTAATTGTAGAGCGGTATTTGTACGGGGTTTTTTGGGG 3576

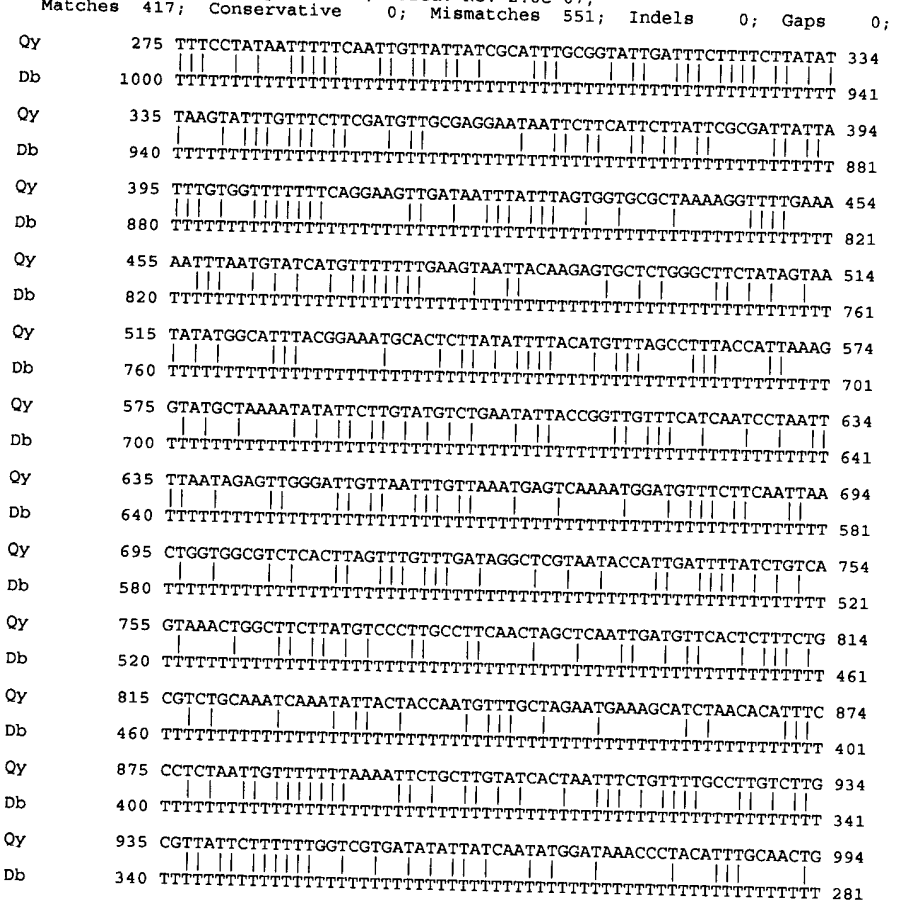
Qy       1175 TATCTA--TGGTAAAAATAATATATCCGGCTTTTCAATTTTATTACCTTTATGTAGCTTT 1232
              | | ||  |  |   | |      ||       |||  |||| ||  | || | ||  ||
Db       3577 TTTTTAGTTTTTTTTAGTGTAGAATTTAAGGTTTTTATTTTTTTTTCGTTTTTTATTTTA 3636

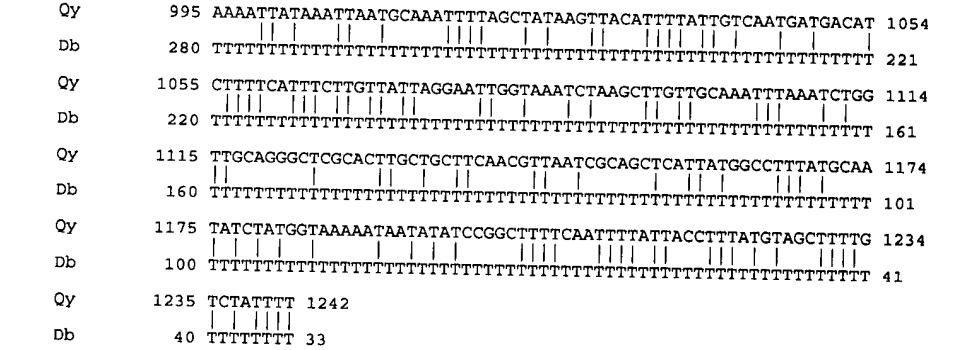
Qy       1233 TGTCTATTTTAAT 1245
              || |||||| ||
Db       3637 CGTTTATTTTTAT 3649
```

RESULT 7
ADQ62833/c
ID ADQ62833 standard; RNA; 1000 BP.
XX
AC ADQ62833;
XX
DT 07-OCT-2004 (first entry)
XX
DE Homopoly-A contaminant for RNaseH activity assay.
XX
KW ss; nuclease-mediated cleavage; target nucleic acid; RNaseH activity;
KW fluorophore; fluorescence quencher; fluorescent signal; HIV;
KW reverse transcriptase.
XX
OS Synthetic.
XX
PN WO2004059012-A1.
XX
PD 15-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US040879.
XX
PR 23-DEC-2002; 2002US-0436125P.
XX
PA (AMHP) WYETH.
XX
PI Olson MW, O'connell JF;
XX
DR WPI; 2004-543471/52.
XX
PT Detecting a nuclease-mediated cleavage of a target nucleic acid, useful
PT for detecting and monitoring RNase H activity, comprises hybridizing a
PT target nucleic acid to a fluorescently labeled oligonucleotide probe.
XX
PS Example 1; SEQ ID NO 4; 61pp; English.
XX
CC The invention relates to a method of detecting a nuclease-mediated
CC cleavage of a target nucleic acid or measuring a RNaseH activity of an
CC agent by hybridizing a target nucleic acid to a fluorescently labelled
CC oligonucleotide probe complementary to the target nucleic acid and
CC containing a fluorophore at one terminus and a quenching group at the
CC other terminus and contacting the probe-target hybrid with an agent
CC having nuclease activity. When the oligonucleotide probe is unhybridised

CC to the target nucleic acid, the probe adopts a conformation that places
CC the fluorophore and quencher in such proximity that the quencher quenches
CC the fluorescent signal of the fluorophore. Formation of the probe-target
CC hybrid causes sufficient separation of the fluorophore and quencher to
CC reduce quenching of the fluorescent signal of the fluorophore followed by
CC contacting the probe-target hybrid with an agent having nuclease activity
CC to selectively cleave the target nucleic acid and thus release the intact
CC probe. Detection of the release of the probe is by measuring a decrease
CC in the fluorescent signal of the fluorophore as compared to the signal of
CC the probe-target hybrid. The methods are useful for detecting and
CC monitoring RNaseH activity in real time. The methods are also useful for
CC monitoring enzymatic degradation of an RNA-DNA duplex. In an example of
CC the invention, the activity of the RNaseH from E. coli or HIV reverse
CC transcriptase is measure by the method of the invention. in order to test
CC the specificity of the RNaseH cleavage of the target-probe complex, a DNA
CC or RNA contaminant is added to the assay reaction. This sequence
CC corresponds to the single stranded homopolymeric polyA RNA contaminant
CC for testing this activity.
XX
SQ Sequence 1000 BP; 1000 A; 0 C; 0 G; 0 T; 0 U; 0 Other;

Query Match 6.8%; Score 86.4; DB 12; Length 1000;
Best Local Similarity 43.1%; Pred. No. 2.6e-07;
Matches 417; Conservative 0; Mismatches 551; Indels 0; Gaps 0;





RESULT 8
ADQ62832
ID ADQ62832 standard; RNA; 1000 BP.
XX
AC ADQ62832;
XX
DT 07-OCT-2004 (first entry)
XX
DE Homopoly-U contaminant for RNaseH activity assay.
XX
KW ss; nuclease-mediated cleavage; target nucleic acid; RNaseH activity;
KW fluorophore; fluorescence quencher; fluorescent signal; HIV;
KW reverse transcriptase.
XX
OS Synthetic.
XX
PN WO2004059012-A1.
XX
PD 15-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US040879.
XX
PR 23-DEC-2002; 2002US-0436125P.
XX
PA (AMHP) WYETH.
XX
PI Olson MW, O'connell JF;
XX
DR WPI; 2004-543471/52.
XX
PT Detecting a nuclease-mediated cleavage of a target nucleic acid, useful
PT for detecting and monitoring RNase H activity, comprises hybridizing a
PT target nucleic acid to a fluorescently labeled oligonucleotide probe.
XX
PS Example 1; SEQ ID NO 3; 61pp; English.
XX
CC The invention relates to a method of detecting a nuclease-mediated
CC cleavage of a target nucleic acid or measuring a RNaseH activity of an
CC agent by hybridizing a target nucleic acid to a fluorescently labelled
CC oligonucleotide probe complementary to the target nucleic acid and
CC containing a fluorophore at one terminus and a quenching group at the
CC other terminus and contacting the probe-target hybrid with an agent
CC having nuclease activity. When the oligonucleotide probe is unhybridised
CC to the target nucleic acid, the probe adopts a conformation that places
CC the fluorophore and quencher in such proximity that the quencher quenches
CC the fluorescent signal of the fluorophore. Formation of the probe-target
CC hybrid causes sufficient separation of the fluorophore and quencher to
CC reduce quenching of the fluorescent signal of the fluorophore followed by
CC contacting the probe-target hybrid with an agent having nuclease activity
CC to selectively cleave the target nucleic acid and thus release the intact
CC probe. Detection of the release of the probe is by measuring a decrease
CC in the fluorescent signal of the fluorophore as compared to the signal of

```
CC   the probe-target hybrid. The methods are useful for detecting and
CC   monitoring RNaseH activity in real time. The methods are also useful for
CC   monitoring enzymatic degradation of an RNA-DNA duplex. In an example of
CC   the invention, the activity of the RNaseH from E. coli or HIV reverse
CC   transcriptase is measure by the method of the invention. in order to test
CC   the specificity of the RNaseH cleavage of the target-probe complex, a DNA
CC   or RNA contaminant is added to the assay reaction. This sequence
CC   corresponds to the single stranded homopolymeric polyU RNA contaminant
CC   for testing this activity.
XX
SQ   Sequence 1000 BP; 0 A; 0 C; 0 G; 0 T; 1000 U; 0 Other;

  Query Match             6.8%;  Score 86.4;  DB 12;  Length 1000;
  Best Local Similarity   0.0%;  Pred. No. 2.6e-07;
  Matches    0;  Conservative  417;  Mismatches  551;  Indels    0;  Gaps    0;

Qy        275 TTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATAT 334
              :::  : :  ::::::   :: :: :: :    :::    : ::  ::: :::: :: : :
Db          1 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 60

Qy        335 TAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTA 394
              :   : ::: ::: ::   : ::         :  :: ::  :: :: ::     :: ::
Db         61 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 120

Qy        395 TTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAA 454
              ::: :  :::::::      ::  :  ::: :::  :  :   :     ::::
Db        121 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 180

Qy        455 AATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAA 514
                :::  : : :  : :::::::    :  ::       :  : :    :: : :  :
Db        181 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 240

Qy        515 TATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAG 574
              : : :    :::       :    : :: : ::::   : :::    :::    ::
Db        241 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 300

Qy        575 GTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCATCAATCCTAATT 634
              : :  :    : : :: :: : : :  : ::     :: :::  :   :  :  ::
Db        301 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 360

Qy        635 TTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTTTCTTCAATTAA 694
              ::  :    ::    :: ::  ::: ::   :   :      :  : ::: ::   ::
Db        361 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 420

Qy        695 CTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGATTTTATCTGTCA 754
              :  :    : :   ::  ::: :::  :    :  :  :    ::  :::: : : :
Db        421 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 480

Qy        755 GTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATGTTCACTCTTTCTG 814
              :   :   :: :: : :   ::   ::   :   :   ::  : ::   : ::: :
Db        481 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 540

Qy        815 CGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCATCTAACACATTTC 874
              : :      :    : ::  :      : :::  :    :       : :       :::
Db        541 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 600

Qy        875 CCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTTTTGCCTTGTCTTG 934
              : :  :: :::::::    :: :  :: : :   :  ::: : ::::   :: : ::
Db        601 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 660

Qy        935 CGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATAAACCCTACATTTGCAACTG 994
              :: :: ::::::  :  :  : : :: :  : :   :       :  :::     :
Db        661 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 720

Qy        995 AAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTTTATTGTCAATGATGACAT 1054
              :: :   ::  :     ::::   : :   ::  :::: :: :   :  :    :
Db        721 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 780

Qy       1055 CTTTTCATTTCTTGTTATTAGGAATTGGTAAATCTAAGCTTGTTGCAAATTTAAATCTGG 1114
               ::::  ::: :: :: ::     ::  :   : :    :: ::     :::   : :
Db        781 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 840

Qy       1115 TTGCAGGGCTCGCACTTGCTGCTTCAACGTTAATCGCAGCTCATTATGGCCTTTATGCAA 1174
                :  ::      :     ::  :  ::     ::  :      :  :: :    ::: :
Db        841 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 900

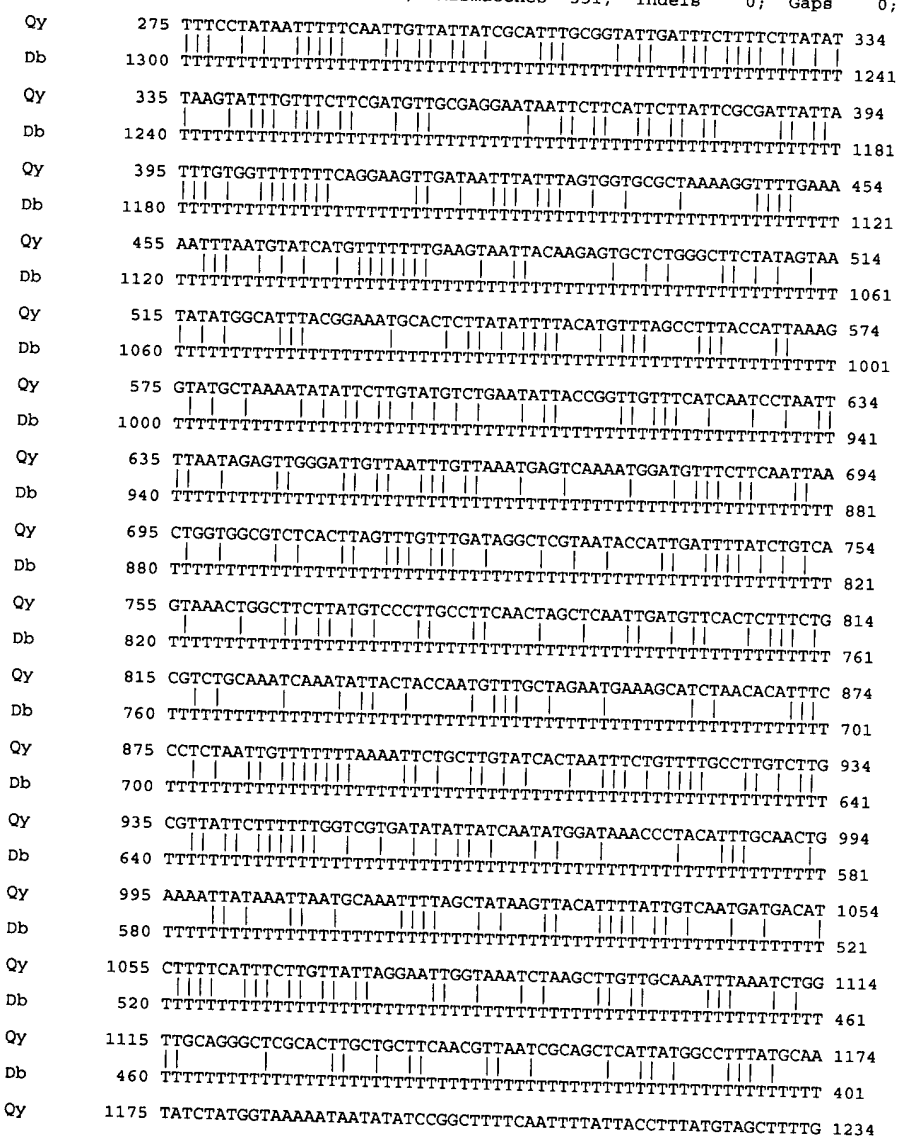
Qy       1175 TATCTATGGTAAAAATAATATATCCGGCTTTTCAATTTTATTACCTTTATGTAGCTTTTG 1234
              : : :  :     :  : : :     ::::   :::: ::   ::: : :   ::::
Db        901 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 960

Qy       1235 TCTATTTT 1242
              : : ::::
Db        961 UUUUUUUU 968


RESULT 9
ADP85917/c
ID   ADP85917 standard; DNA; 1300 BP.
XX
AC   ADP85917;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Synthetic construct #1.
XX
KW   Nanopore data analysis ; polymer ; polynucleotide;
KW   single-molecule resolution; ds.
XX
OS   Synthetic.
XX
PN   US2004110205-A1.
XX
PD   10-JUN-2004.
XX
PF   23-SEP-2003; 2003US-00668749.
XX
PR   23-SEP-2002; 2002US-0412959P.
XX
PA   (WANG/) WANG H.
XX
PI   Wang H;
XX
DR   WPI; 2004-467656/44.
XX
PT   Nanopore analysis system for analyzing polynucleotides includes nanopore
PT   device having structure with aperture, and nanopore data analysis system
PT   that can generate nanopore data points corresponding to each target and
PT   non-target polymer.
XX
PS   Disclosure; SEQ ID NO 1; 17pp; English.
XX
CC   The invention relates to a nanopore analysis system that comprises a
CC   nanopore device and a nanopore data analysis system. The nanopore device
CC   has a structure having an aperture. The nanopore data analysis system can
CC   generate nanopore data points corresponding to each target polymer and
CC   each non-target polymer traversing the aperture; form a distribution
CC   pattern of the data points; and analyse a distribution of target polymer
CC   data points in the distribution pattern. The distribution pattern
CC   includes data cluster(s). The nanopore data analysis system is stored on
CC   a computer-readable medium. A mechanism is provided for analysing the
CC   distribution of target polynucleotide data points in the distribution
CC   pattern. The nanopore data analysis system can also analyse the
CC   distribution of the non-target polynucleotide data points. It can also
CC   determine a ratio between the target polynucleotide data points and the
CC   non-target polynucleotide data points. It can further analyse the
CC   distribution of target polynucleotide data points between the two data
CC   clusters; compare the distribution of the target polynucleotide data
CC   points between the two data clusters to a phosphorylation state standard
CC   distribution; and determine a ratio of phosphorylated target
CC   polynucleotide to non-phosphorylated target polynucleotides. It can also
CC   determine a cluster score for the target polynucleotide data points in a
CC   defined area; and compare the cluster score for the target polynucleotide
CC   data points to a cluster score for a chemical integrity standard density
CC   distribution for the defined area in a distribution of a target
CC   polynucleotide standard. The invention is useful for analysing polymers,
CC   such as polynucleotides. The inventive nanopore analysis system
```

CC potentially provides high speed sampling with single-molecule resolution,
CC which may enable unprecedented dynamic range and sensitivity in analysis
CC of samples containing charged polymers such as polynucleotides or
CC polypeptides. The present sequence is a synthetic construct used in the
CC invention.
XX
SQ Sequence 1300 BP; 1300 A; 0 C; 0 G; 0 T; 0 U; 0 Other;

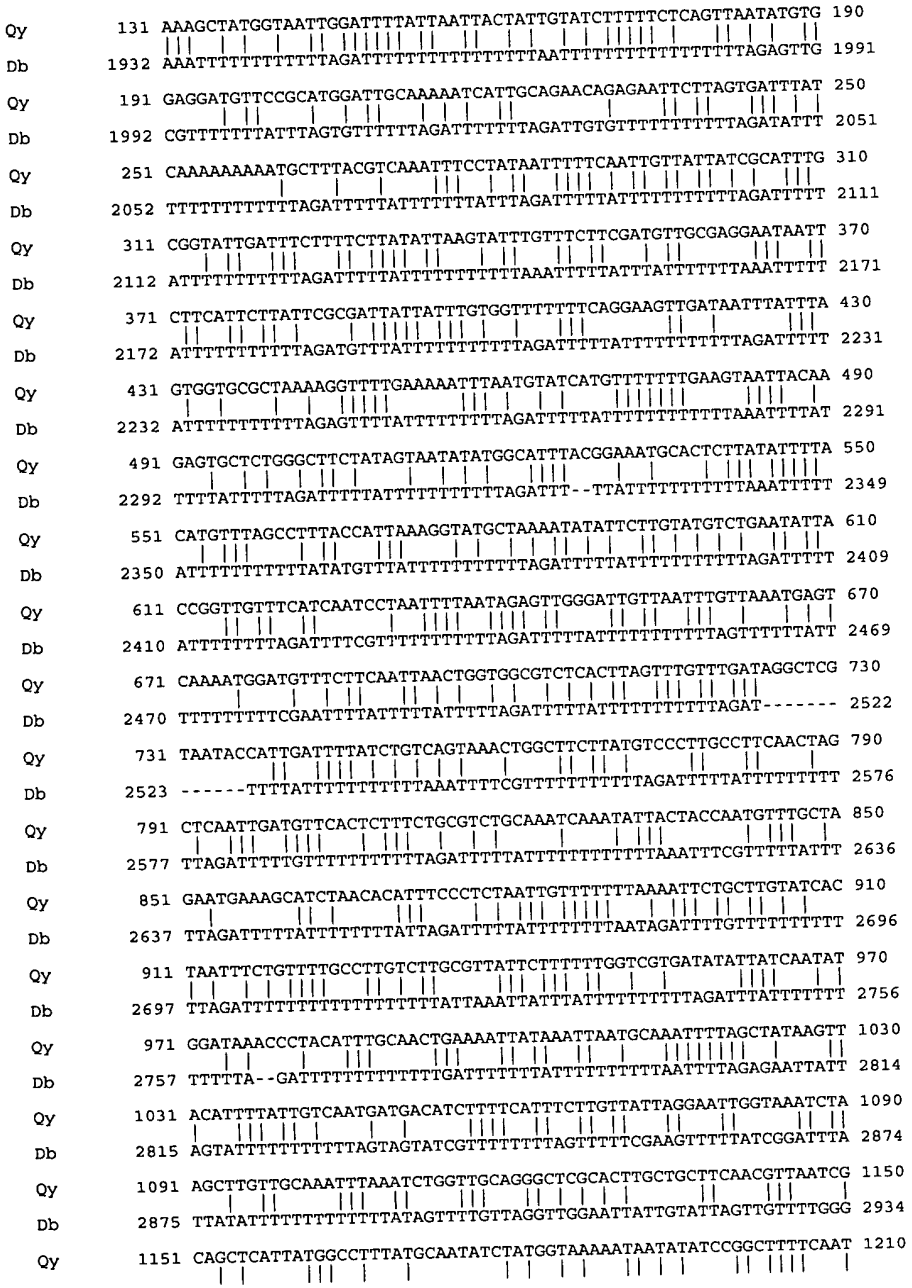
Query Match 6.8%; Score 86.4; DB 12; Length 1300;
Best Local Similarity 43.1%; Pred. No. 2.6e-07;
Matches 417; Conservative 0; Mismatches 551; Indels 0; Gaps 0;

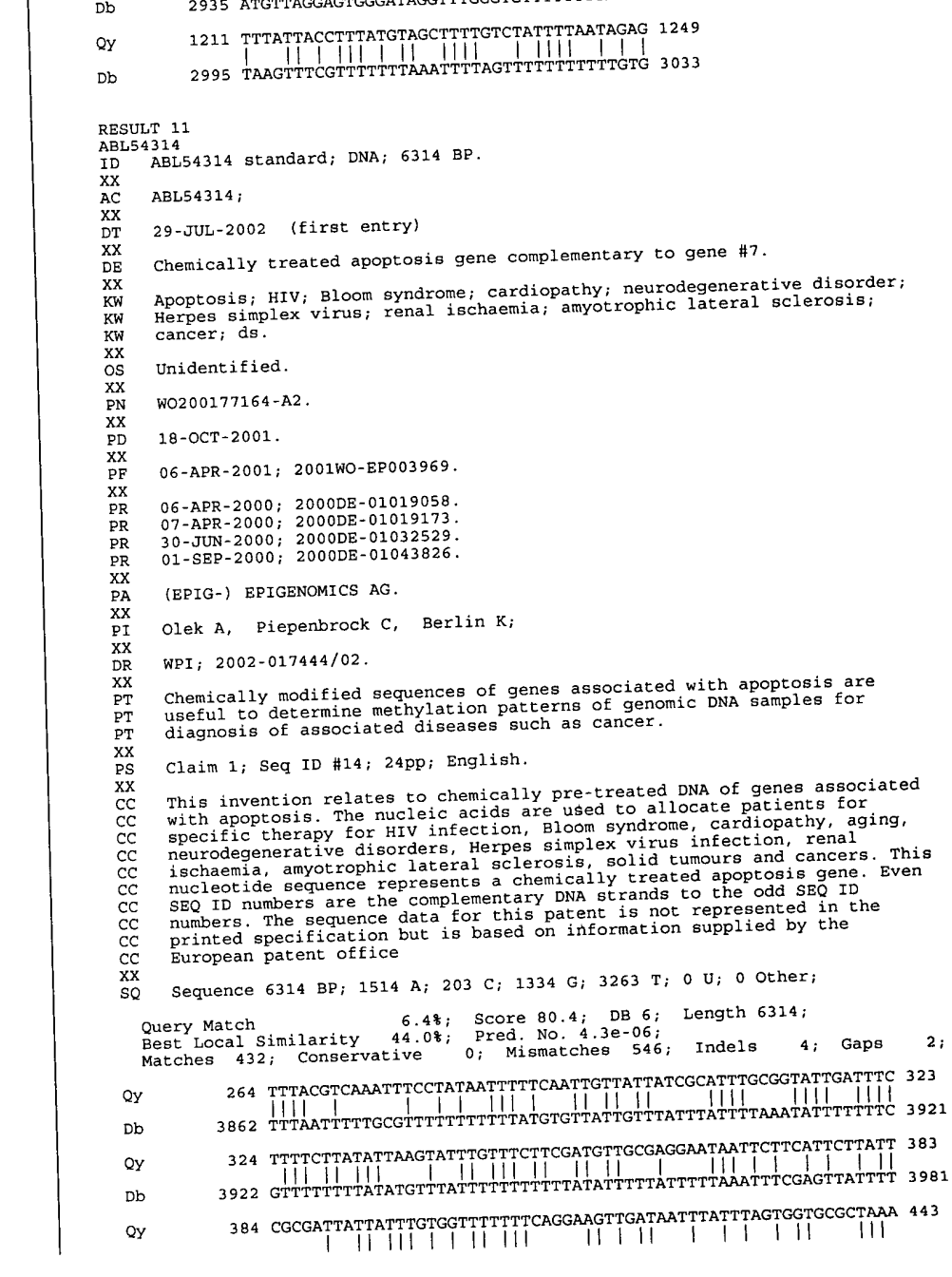




RESULT 10
AAS46686
ID AAS46686 standard; DNA; 7442 BP.
XX
AC AAS46686;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #409.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 409; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 7442 BP; 1655 A; 214 C; 1934 G; 3636 T; 0 U; 3 Other;

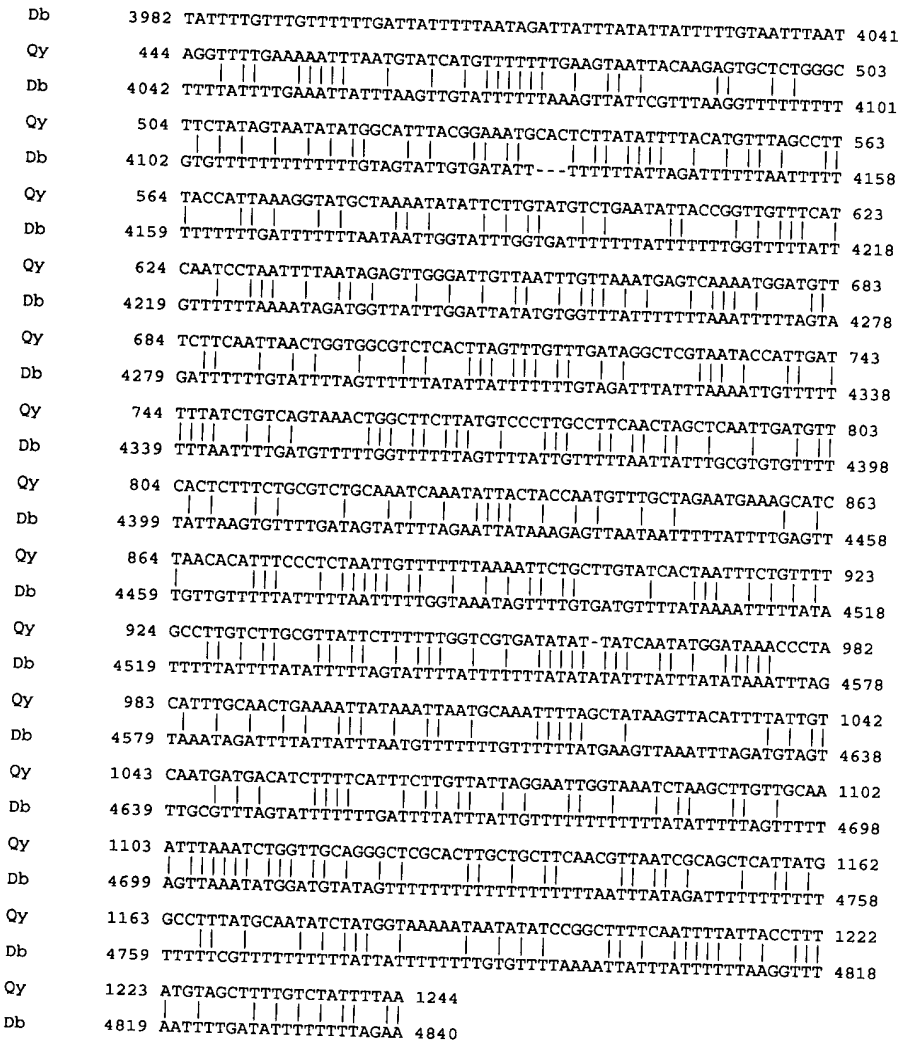
Query Match 6.5%; Score 82.2; DB 4; Length 7442;





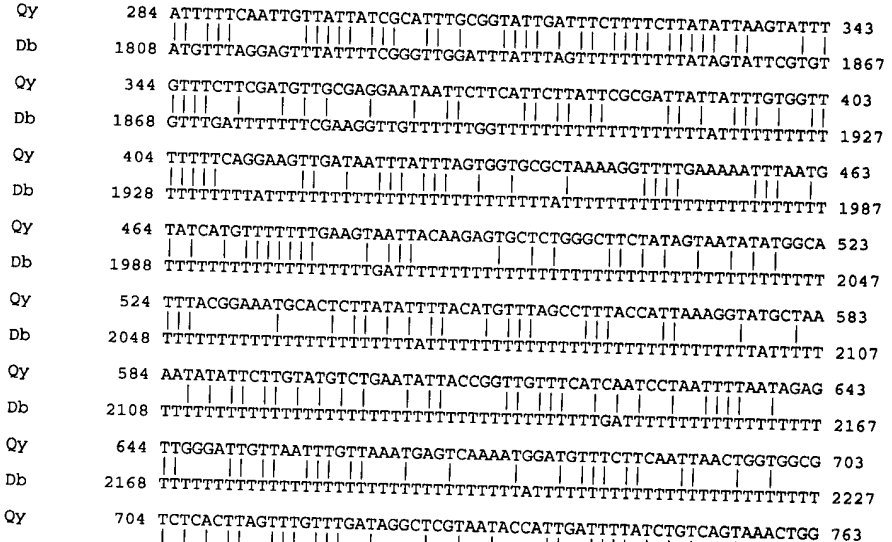
RESULT 11
ABL54314
ID ABL54314 standard; DNA; 6314 BP.
XX
AC ABL54314;
XX
DT 29-JUL-2002 (first entry)
XX
DE Chemically treated apoptosis gene complementary to gene #7.
XX
KW Apoptosis; HIV; Bloom syndrome; cardiopathy; neurodegenerative disorder;
KW Herpes simplex virus; renal ischaemia; amyotrophic lateral sclerosis;
KW cancer; ds.
XX
OS Unidentified.
XX
PN WO200177164-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP003969.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017444/02.
XX
PT Chemically modified sequences of genes associated with apoptosis are
PT useful to determine methylation patterns of genomic DNA samples for
PT diagnosis of associated diseases such as cancer.
XX
PS Claim 1; Seq ID #14; 24pp; English.
XX
CC This invention relates to chemically pre-treated DNA of genes associated
CC with apoptosis. The nucleic acids are used to allocate patients for
CC specific therapy for HIV infection, Bloom syndrome, cardiopathy, aging,
CC neurodegenerative disorders, Herpes simplex virus infection, renal
CC ischaemia, amyotrophic lateral sclerosis, solid tumours and cancers. This
CC nucleotide sequence represents a chemically treated apoptosis gene. Even
CC SEQ ID numbers are the complementary DNA strands to the odd SEQ ID
CC numbers. The sequence data for this patent is not represented in the
CC printed specification but is based on information supplied by the
CC European patent office
XX
SQ Sequence 6314 BP; 1514 A; 203 C; 1334 G; 3263 T; 0 U; 0 Other;

Query Match 6.4%; Score 80.4; DB 6; Length 6314;
Best Local Similarity 44.0%; Pred. No. 4.3e-06;
Matches 432; Conservative 0; Mismatches 546; Indels 4; Gaps 2;



RESULT 12
ABL33958
ID ABL33958 standard; DNA; 14006 BP.
XX
AC ABL33958;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1931.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1931; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 14006 BP; 3278 A; 155 C; 3257 G; 7313 T; 0 U; 3 Other;

Query Match 6.2%; Score 78; DB 6; Length 14006;
Best Local Similarity 43.8%; Pred. No. 1.3e-05;
Matches 339; Conservative 0; Mismatches 435; Indels 0; Gaps 0;



```
Db       2228 TTTTATTTTTTTTTTTTCGTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTT 2287

Qy        764 CTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAA 823

Db       2288 TTTTTTTTTTTTCGTTTTTTTTTTTTTTTTTTTTTTTATATTTTTTTTTTTTTTTTTT 2347

Qy        824 ATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATT 883

Db       2348 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTGGTTTTTTTTTTTTTTTTTTTTTTT 2407

Qy        884 GTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCT 943

Db       2408 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2467

Qy        944 TTTTTGGTCGTGATATATTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATA 1003

Db       2468 TTTTTTTTTTTTTTTTTTTTTTTTTATTGTTTATAGTAGATGAGGAAATTGAGATTTTAA 2527

Qy       1004 AATTAATGCAAATTTTAGCTATAAGTTACATTTTATTGTCAATGATGACATCTT 1057

Db       2528 AGGTGAAAGGTTTTGTTTAAGTTATATAGTTTATAGTGATAGCGATAGGAATTT 2581
```

RESULT 13
ACL64817/c
ID ACL64817 standard; DNA; 1039 BP.
XX
AC ACL64817;
XX
DT 02-JUN-2005 (first entry)
XX
DE M. xanthus DNA fragment, seq id 1280.
XX
KW Transgenic plant; DNA replication; gene regulation; gene expression; ds.
XX
OS Myxococcus xanthus.
XX
PN US6833447-B1.
XX
PD 21-DEC-2004.
XX
PF 10-JUL-2001; 2001US-00902540.
XX
PR 10-JUL-2000; 2000US-0217883P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX
DR WPI; 2005-028716/03.
XX
PT New substantially purified Myxococcus xanthus nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX
PS Example 1; SEQ ID NO 1280; 25pp; English.
XX
CC The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC set of 1849 contig and singleton sequences comprising coding sequences,
CC DNA replication elements, promoters and other regulatory elements from
CC the genome of the bacterium Mxyococcus xanthus. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO
XX
SQ Sequence 1039 BP; 819 A; 72 C; 17 G; 129 T; 0 U; 2 Other;

Query Match 6.1%; Score 77.4; DB 14; Length 1039;
Best Local Similarity 45.6%; Pred. No. 1.4e-05;
Matches 425; Conservative 0; Mismatches 498; Indels 10; Gaps 4;

```
Qy        144 TTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTGGAGGATGTTCCGC 203

Db        966 TTTTTTTTTTTTATGGTTTTATTTTTTTTTTTTTGGATATTTTGTAATTTTTTTTTTC 907

Qy        204 ATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGATTTATCAAAAAAAAATGC 263

Db        906 TTTTATTTTTTTATTTTTTATTTGGTTTTTTATTTTTTTTAAGTTATTTTTGTTTTTTAT 847

Qy        264 TTTACGTCAAATTTCCTATAATT----TTTCAATTGTTATTATCGCATTTGCGGTATTGA 319

Db        846 TTTCCCTCATTTTTTTTAGTGTTTTTAATTTTTTTTTTTTTTTAGAAGGTTTATATTTTTAT 787

Qy        320 TTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCT 379

Db        786 ATTTTTTTTTTGTTTTTAGGATTTTTTGTTATATTGGGGTGTTGATTAGAGC--CTTTGT 729

Qy        380 TATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGC 439

Db        728 TAATTTCTTTTTTTTTTTTTTTATGTTTTTAGGGGAGATTTTTTTTTTTTTAAGAAAAAA 669

Qy        440 TAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCT 499

Db        668 CACTTTTTTTTGTTTTTTTATAGTTTTATTTATTTTTATATTTTATGCGGNGTTTTTTTT 609

Qy        500 GGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAG 559

Db        608 TTGTTTTTTAAATTATGTATTTTACTTTTTTTTTTTTTTTATCTTTTTTTATTTTTTTTT 549

Qy        560 CCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTT 619

Db        548 TATTTGTTTTTTTTTTTTTTATTTTTTTATTTTTATTTTTTTGTTTTCATTTTTTTTTTTT 489

Qy        620 TCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGA 679

Db        488 TTATTTATTTTTTTTTTTTTGGTGTTTTTGTTGTTTTTTTTTTACAT--ATTATTTTTTT 431

Qy        680 TGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCAT 739

Db        430 TTTTTTTTTANTTTTCTTTTATTTTGTTGTAAATTAATTTTTTTATTTTAATTTTTTGTT 371

Qy        740 TGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGA 799

Db        370 TTTTTTGTTATTTATTTAATTTTTTTTATTTTATGATTGATTTTAATTTTTTTTTTTTTT 311

Qy        800 TGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAG 859

Db        310 ATTTTTTTTTTTTTTTGATTTTTTTTATTTATTTTTTTTTTTTTTTTTTTTTATGTATTGTTT 251

Qy        860 CATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTG 919

Db        250 TTTTTTATATATTTTTTTTTTTTTTTTTTTTTATAATGTTTTTTTTTTTTTTTTTTTTTTT 191

Qy        920 TTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATAAACC 979

Db        190 TTTTTTTTTTATTTTTTTTTTTTAATATTTTTTTTTTTTTTTTTTTTATATTTGTTTTTATT 131

Qy        980 CTACATTTGCAAC--TGAAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTTT 1037

Db        130 TTTTTTTTATTAATTTGTTTTTTTTATTTTTTTTTTTTTTTATTTTTTGTTTTGTTTTT 71

Qy       1038 ATTGTCAATGATGACATCTTTTCATTTCTTGTT 1070

Db         70 AAGTTTTTTGAAGGCTTATTTTTTTTATTTTTT 38
```

RESULT 14
AAX33181/c
ID AAX33181 standard; DNA; 6644 BP.
XX

```
AC   AAX33181;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Base sequence of the plasmid pRx-ires-bsr.
XX
KW   Cowpox virus; bsr; viral vector; expression; apoptosis; resistance; crmA;
KW   bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW   autoimmune disease; graft rejection reaction; inflammation;
KW   inflammatory disease; ss.
XX
OS   Synthetic.
OS   Cowpox virus.
XX
PN   WO9913073-A2.
XX
PD   18-MAR-1999.
XX
PF   07-SEP-1998;   98WO-JP004010.
XX
PR   08-SEP-1997;   97JP-00259235.
XX
PA   (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI   Hamada H;
XX
DR   WPI; 1999-243728/20.
XX
PT   New apoptosis-resistant virus-sensitive cell.
XX
PS   Example 1; Page 38-41; 51pp; English.
XX
CC   The present invention describes an apoptosis-resistant virus-sensitive
CC   cell line into which an apoptosis resistance gene has been introduced.
CC   The recombinant viruses generated are capable of expressing apoptosis-
CC   associated genes. These can then be used in a variety of diseases for
CC   which the induction of apoptosis by gene transfer, or where the
CC   inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC   are useful as vectors for gene therapy which can be applied to cancer
CC   therapy for destroying cancer cells selectively, the treatment of
CC   autoimmune diseases and graft rejection reaction, and apoptosis induction
CC   therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC   encountered the problem where if an adenovirus vector capable of
CC   expressing an apoptosis-associated gene is introduced into animal cells,
CC   the cells producing the virus will be destroyed because the period of
CC   time required to induce cell death by apoptosis is shorter than that
CC   required to replicate and produce the virus, resulting in failure to
CC   obtain a recombinant virus having the integrated apoptosis-associated
CC   gene. In this invention an apoptosis-resistant 293 cell line (having an
CC   apoptosis resistant gene introduced) is established and overcomes the
CC   problem. The present sequence represents the base sequence of the plasmid
CC   pRx-ires-bsr, which contains the cowpox virus bsr gene, and is used in an
CC   example from the present invention
XX
SQ   Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 U; 0 Other;

  Query Match             6.1%;  Score 76.6;  DB 2;  Length 6644;
  Best Local Similarity   44.3%;  Pred. No. 2.3e-05;
  Matches  310;  Conservative    0;  Mismatches  389;  Indels    0;  Gaps    0;

Qy       266 TACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTT 325
Db      4439 TTCGACCAATTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4380

Qy       326 TTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCG 385
Db      4379 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4320

Qy       386 CGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAG 445
Db      4319 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4260

Qy       446 GTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTT 505
```

```
Db      4259 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4200

Qy       506 CTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTTTA 565
Db      4199 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4140

Qy       566 CCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCATCA 625
Db      4139 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4080

Qy       626 ATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTTTC 685
Db      4079 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4020

Qy       686 TTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGATTT 745
Db      4019 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3960

Qy       746 TATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATGTTCA 805
Db      3959 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3900

Qy       806 CTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCATCTA 865
Db      3899 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3840

Qy       866 ACACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTTTTGC 925
Db      3839 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3780

Qy       926 CTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATATTAT 964
Db      3779 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3741
```

```
RESULT 15
AAX33182/c
ID   AAX33182 standard; DNA; 7372 BP.
XX
AC   AAX33182;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Base sequence of the plasmid pRx-Bcl-xl-bsr.
XX
KW   Cowpox virus; bsr; viral vector; expression; apoptosis; resistance; crmA;
KW   bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW   autoimmune disease; graft rejection reaction; inflammation;
KW   inflammatory disease; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9913073-A2.
XX
PD   18-MAR-1999.
XX
PF   07-SEP-1998;   98WO-JP004010.
XX
PR   08-SEP-1997;   97JP-00259235.
XX
PA   (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI   Hamada H;
XX
DR   WPI; 1999-243728/20.
XX
PT   New apoptosis-resistant virus-sensitive cell.
XX
PS   Example 2; Page 41-45; 51pp; English.
XX
CC   The present invention describes an apoptosis-resistant virus-sensitive
```

CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the plasmid
CC pRx-Bcl-xl-bsr, which contains the human Bcl-xl gene, and is used in an
CC example from the present invention
XX
SQ Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T; 0 U; 0 Other;

Query Match 6.1%; Score 76.6; DB 2; Length 7372;
Best Local Similarity 44.3%; Pred. No. 2.3e-05;
Matches 310; Conservative 0; Mismatches 389; Indels 0; Gaps 0;

```
Qy        266 TACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTT 325
              | || | ||||   | |  |||||  || || || |   |||    | ||  ||| ||
Db       5167 TTCGACCAATTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5108

Qy        326 TTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCG 385
              || || | ||   | ||| ||| ||   | ||         |  || ||  || || ||
Db       5107 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5048

Qy        386 CGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAG 445
                || || ||| |  |||||||        ||  |  ||| |||  |  |    |
Db       5047 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4988

Qy        446 GTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTT 505
               ||||      |||  | | |  | |||||||    |  ||       |  | |    ||
Db       4987 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4928

Qy        506 CTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTTTA 565
               | |  |  | | |    |||        |    | || | ||||   | |||    |||
Db       4927 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4868

Qy        566 CCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCATCA 625
                ||      | |  |     | | || || | | |  |   | ||     || |||  |
Db       4867 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4808

Qy        626 ATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTTTC 685
               |  |  ||||  |    ||    || ||  ||| ||   |    |    |  | |||
Db       4807 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4748

Qy        686 TTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGATTT 745
              ||   ||   |  |    | |   ||  ||| |||  |    |  |  |    ||  |||
Db       4747 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4688

Qy        746 TATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATGTTCA 805
              | | | |    |     |   || || | |   ||   ||    |   |   ||  | ||
Db       4687 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4628

Qy        806 CTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCATCTA 865
               | ||| |   | |    |   | ||  |    | |||  |    |       | |
Db       4627 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4568

Qy        866 ACACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTTTTGC 925
                 |||   | |  || |||||||    || |  || | |   |  ||| | ||||
Db       4567 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4508

Qy        926 CTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATATTAT 964
                || | ||   || || |'||||  |  |  | | || |
```

Db 4507 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 4469

Search completed: May 10, 2006, 10:59:38
Job time : 593.987 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 11:00:02 ; Search time 4054.75 Seconds
(without alignments)
14608.170 Million cell updates/sec

Title: US-09-423-093B-1_COPY_8646_9911
Perfect score: 1266
Sequence: 1 atggtattaacagtgaaaaa.....gagcgaaaaatgtctattga 1266

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	105.2	8.3	1482	10	AG332292	AG332292 Mus muscu
c	2	104.2	8.2	1512	10	CL082685	CL082685 CH216-169
	3	104	8.2	1506	10	AG278469	AG278469 Mus muscu
	4	103.8	8.2	1811	10	CG753732	CG753732 P048-4-G0
c	5	103.6	8.2	1205	10	CL143963	CL143963 ISB1-123O
	6	103.6	8.2	1507	10	AG346189	AG346189 Mus muscu
	7	103.4	8.2	1843	10	AG435185	AG435185 Mus muscu
c	8	102.6	8.1	1632	10	CL082569	CL082569 CH216-167
c	9	102.4	8.1	1243	10	CL016663	CL016663 CH216-1A2
	10	102.4	8.1	1519	10	AG386893	AG386893 Mus muscu
c	11	102.4	8.1	1536	10	CL078538	CL078538 CH216-151
	12	102.4	8.1	1566	10	CG757757	CG757757 P053-1-D0
	13	102.2	8.1	1981	10	CL082000	CL082000 CH216-165
	14	102	8.1	1348	10	CG749499	CG749499 P043-4-A0
	15	101.6	8.0	1472	10	AG350118	AG350118 Mus muscu
c	16	101.4	8.0	1217	10	CL062848	CL062848 CH216-98N
c	17	101.4	8.0	1599	10	CL083840	CL083840 ISB1-2H14
c	18	100.8	8.0	1808	10	CL078963	CL078963 CH216-154
c	19	100.2	7.9	1330	8	DN705530	DN705530 CLJ60-A10
	20	100.2	7.9	1373	10	CG750869	CG750869 P045-2-E1
c	21	100	7.9	1531	10	CG748014	CG748014 P041-4-B0
	22	100	7.9	1559	10	CL646206	CL646206 CH213-110
	23	99.8	7.9	1298	10	CG756607	CG756607 P051-4-C0
	24	99.6	7.9	1297	10	CG758143	CG758143 P053-3-B1
c	25	99.6	7.9	1459	8	DN726869	DN726869 CNB32-G02
	26	99.4	7.9	1376	10	CG747831	CG747831 P041-3-B0
	27	99.4	7.9	1407	10	AJ592026	AJ592026 Arabidops
c	28	99.4	7.9	1468	10	CL078540	CL078540 CH216-151
c	29	99.2	7.8	1163	10	CL081338	CL081338 CH216-161
	30	99.2	7.8	1361	10	CG744327	CG744327 P036-4-E0
c	31	99	7.8	1594	10	CL038406	CL038406 CH216-46A
	32	98.6	7.8	1453	10	AJ591978	AJ591978 Arabidops
	33	98.4	7.8	1220	10	CL641847	CL641847 CH213-19H
	34	98.4	7.8	1248	8	DN808090	DN808090 73970127
	35	98.4	7.8	1275	10	CL647839	CL647839 CH213-152
	36	98.2	7.8	1317	1	AJ926430	AJ926430 AJ926430
c	37	98.2	7.8	1594	10	CL078613	CL078613 CH216-151
	38	98	7.7	1276	10	CG754010	CG754010 P049-2-C0
	39	98	7.7	1491	10	CG753221	CG753221 P048-2-A0
c	40	98	7.7	1596	10	CL081967	CL081967 CH216-165
	41	97.6	7.7	1353	10	CG744812	CG744812 P037-3-B0
	42	97.6	7.7	1434	10	AJ592058	AJ592058 Arabidops
c	43	97.6	7.7	1641	10	CL078595	CL078595 CH216-151
c	44	97.6	7.7	1654	10	CL118709	CL118709 ISB1-72I1
	45	97.4	7.7	1074	9	BZ696936	BZ696936 SP__Ba009

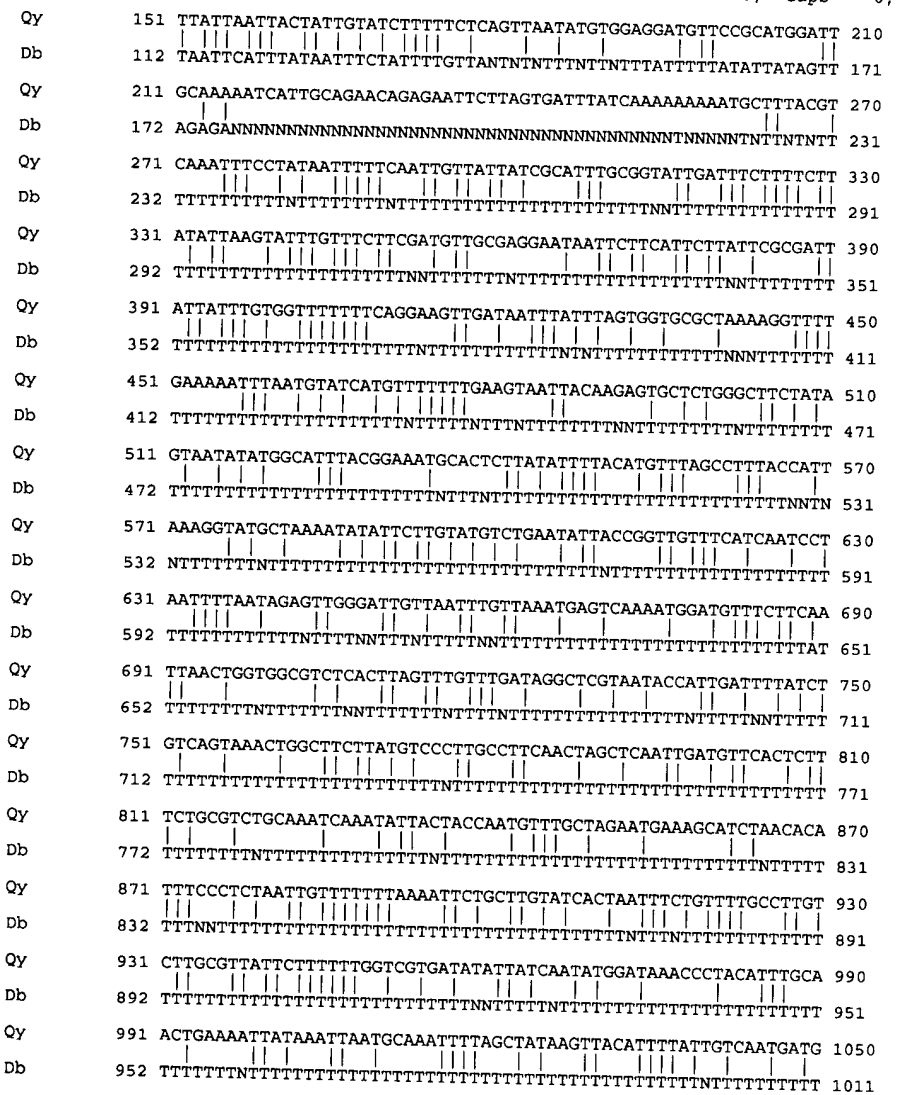
ALIGNMENTS

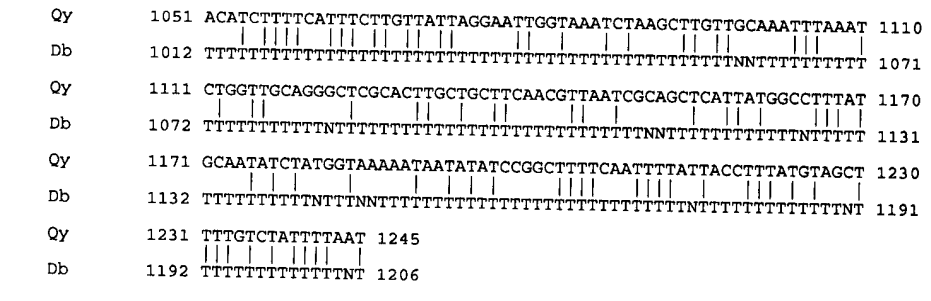
RESULT 1
AG332292
LOCUS AG332292 1482 bp DNA linear GSS 18-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-122G19.TJ, genomic survey sequence.
ACCESSION AG332292
VERSION AG332292.1 GI:47905602
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1482)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1482

/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-122G19.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

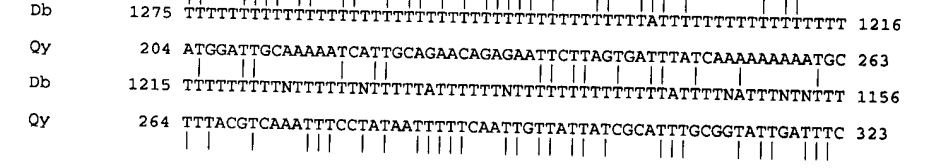
Query Match 8.3%; Score 105.2; DB 10; Length 1482;
Best Local Similarity 39.4%; Pred. No. 2.6e-10;
Matches 431; Conservative 0; Mismatches 664; Indels 0; Gaps 0;

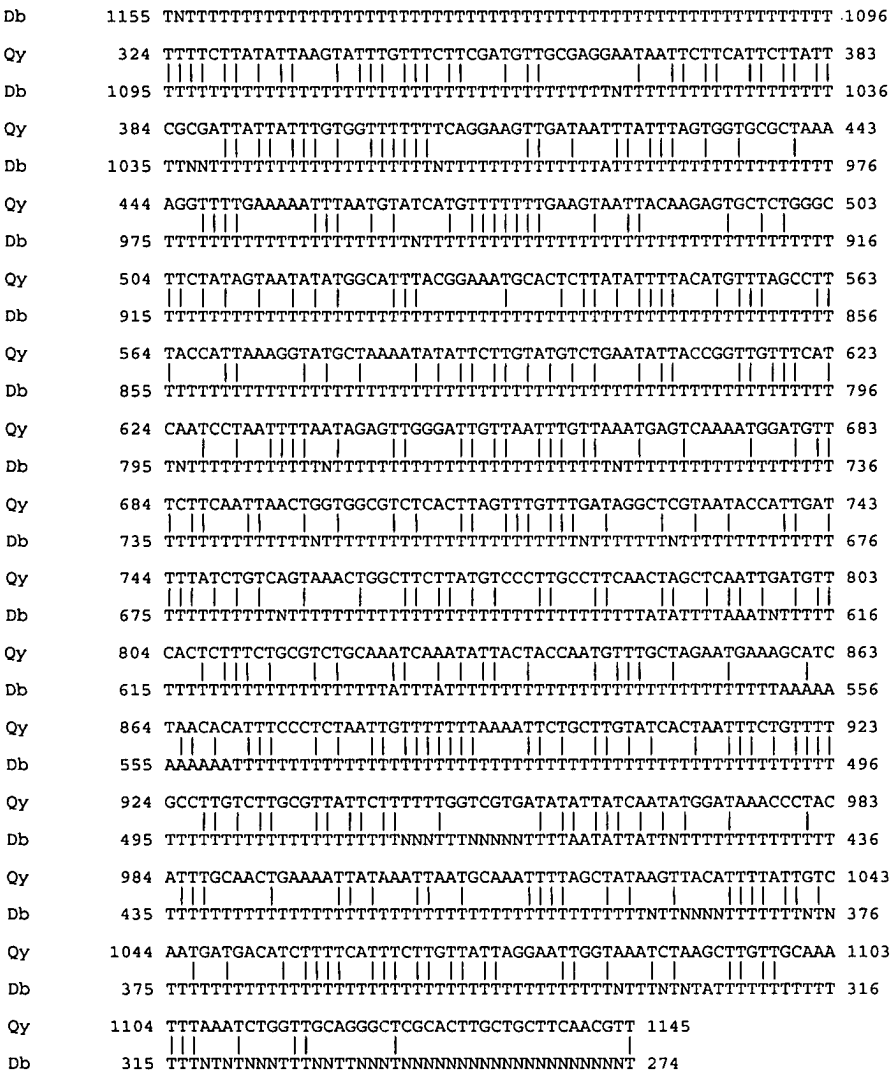




RESULT 2
CL082685/c
LOCUS CL082685 1512 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-169J11_RM4.1 CH216 Xenopus tropicalis genomic clone CH216-169J11, genomic survey sequence.
ACCESSION CL082685
VERSION CL082685.1 GI:40538598
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1512)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM4 ctcaagggcatcggtcgagc
Class: BAC ends
High quality sequence start: 814
High quality sequence stop: 900.
FEATURES Location/Qualifiers
source 1. .1512
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-169J11"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"
ORIGIN

Query Match 8.2%; Score 104.2; DB 10; Length 1512;
Best Local Similarity 41.4%; Pred. No. 4.1e-10;
Matches 415; Conservative 0; Mismatches 587; Indels 0; Gaps 0;

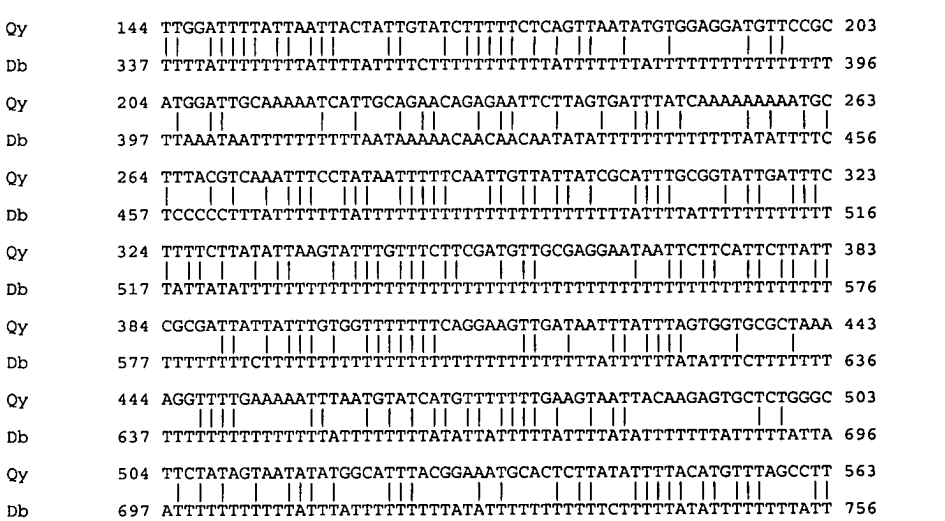


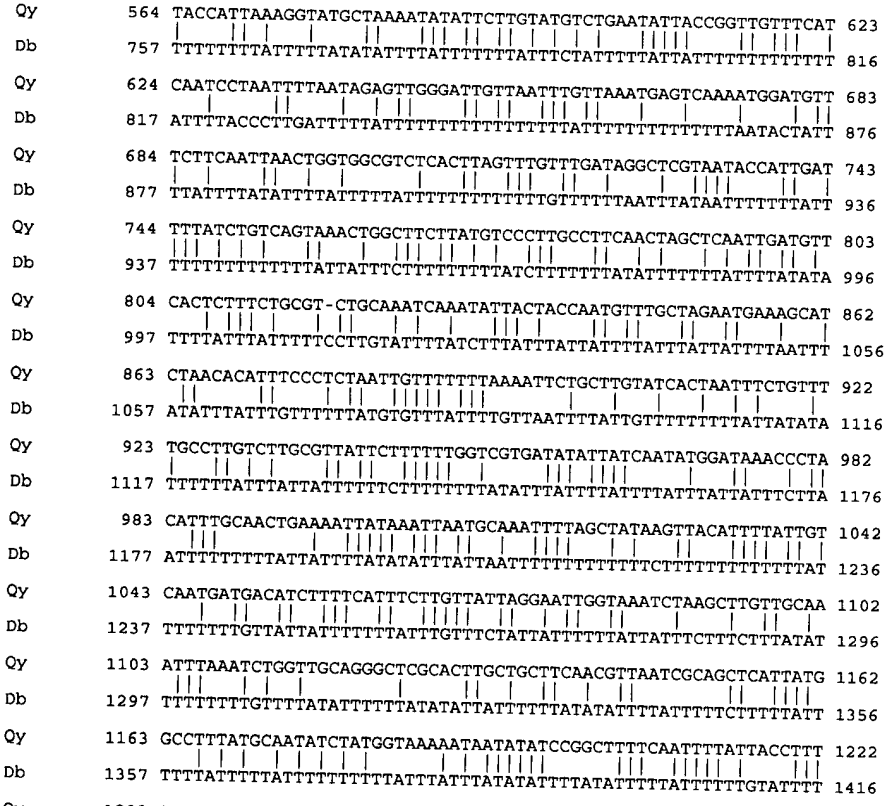


RESULT 3
AG278469
LOCUS AG278469 1506 bp DNA linear GSS 18-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-048M18.T7, genomic survey sequence.
ACCESSION AG278469
VERSION AG278469.1 GI:47851346
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1

AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1506)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1506
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-048M18.T7"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

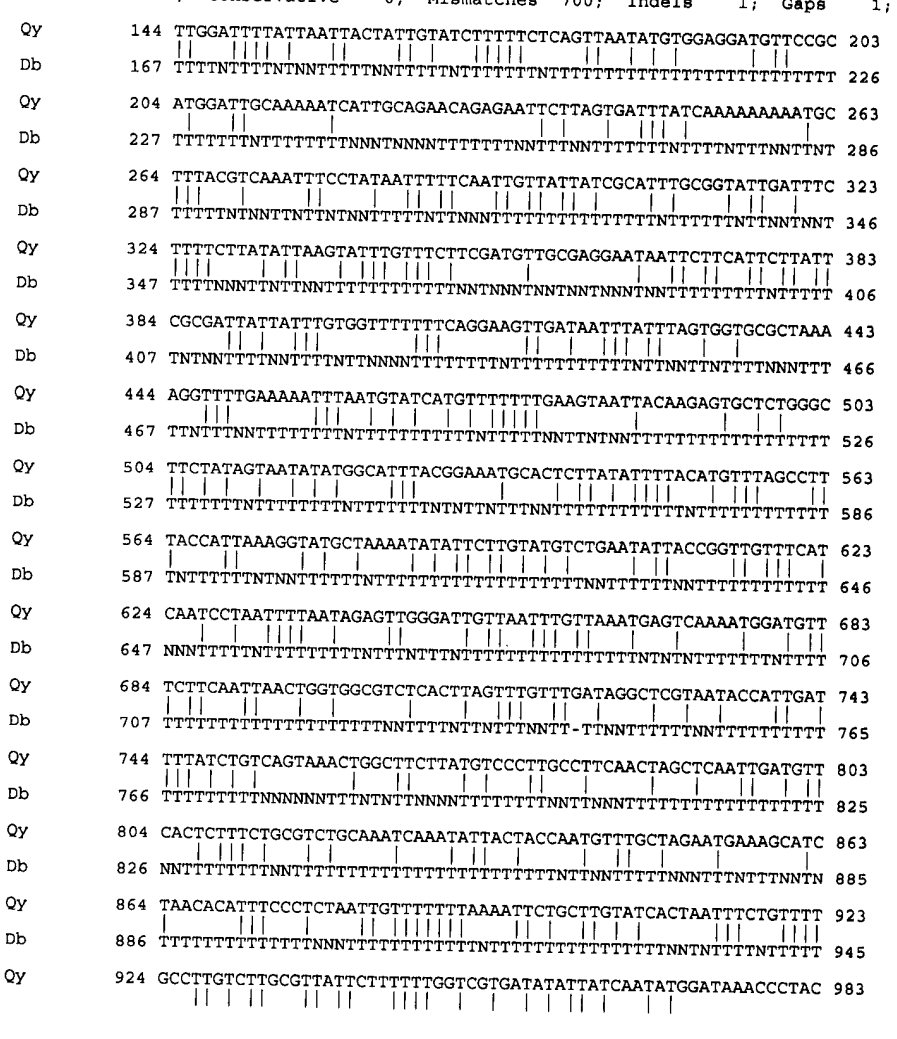
Query Match 8.2%; Score 104; DB 10; Length 1506;
Best Local Similarity 44.0%; Pred. No. 4.5e-10;
Matches 484; Conservative 0; Mismatches 615; Indels 1; Gaps 1;





RESULT 4
CG753732
LOCUS CG753732 1811 bp DNA linear GSS 24-OCT-2003
DEFINITION P048-4-G03.ya Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG753732
VERSION CG753732.1 GI:37978509
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1811)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1811
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."
ORIGIN

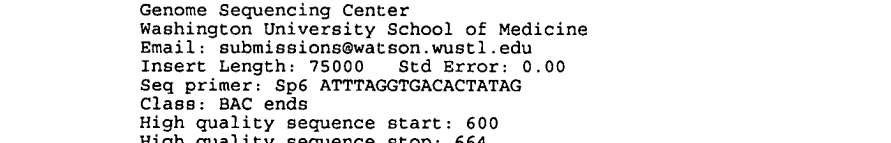
Query Match 8.2%; Score 103.8; DB 10; Length 1811;
Best Local Similarity 36.4%; Pred. No. 4.9e-10;
Matches 401; Conservative 0; Mismatches 700; Indels 1; Gaps 1;





RESULT 5
CL143963/c
LOCUS CL143963 1205 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-123O2_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-123O2, genomic survey sequence.
ACCESSION CL143963
VERSION CL143963.1 GI:40637598
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1205)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 600
High quality sequence stop: 664.
FEATURES Location/Qualifiers
source 1. .1205
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-123O2"
/clone_lib="ISB1"
/note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"
ORIGIN

Query Match 8.2%; Score 103.6; DB 10; Length 1205;
Best Local Similarity 38.6%; Pred. No. 5.4e-10;
Matches 432; Conservative 0; Mismatches 687; Indels 1; Gaps 1;



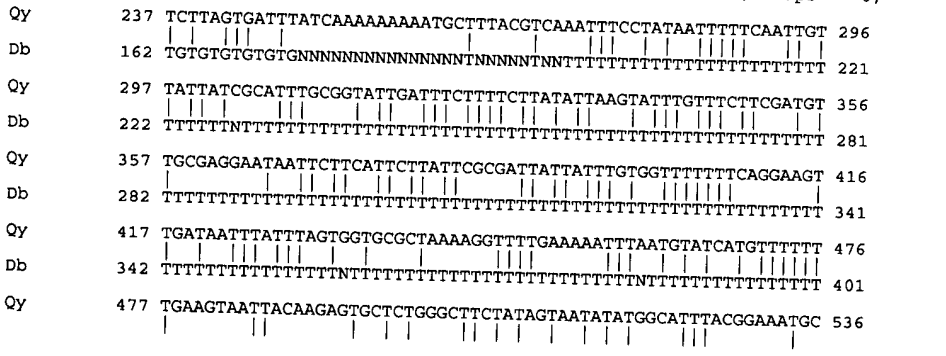


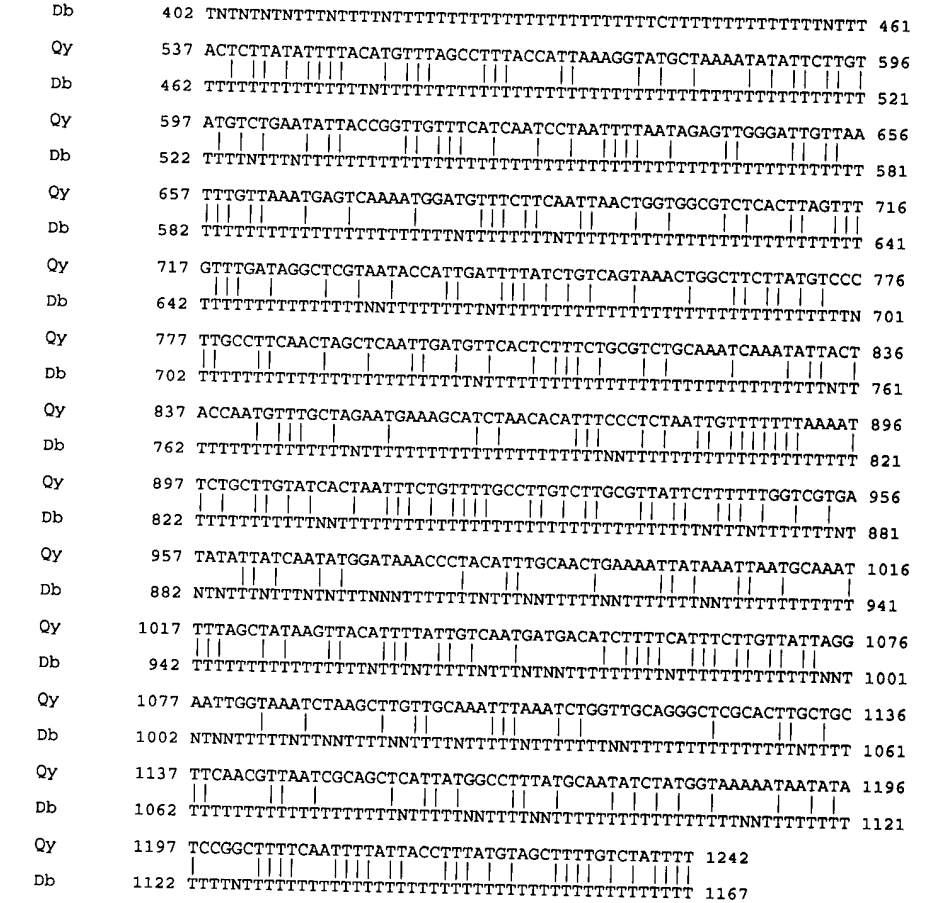


RESULT 6
AG346189
LOCUS AG346189 1507 bp DNA linear GSS 18-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-141E08.TJ, genomic survey

sequence.
ACCESSION AG346189
VERSION AG346189.1 GI:47919499
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1507)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp). Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1507
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-141E08.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Query Match 8.2%; Score 103.6; DB 10; Length 1507;
Best Local Similarity 40.4%; Pred. No. 5.4e-10;
Matches 406; Conservative 0; Mismatches 600; Indels 0; Gaps 0;

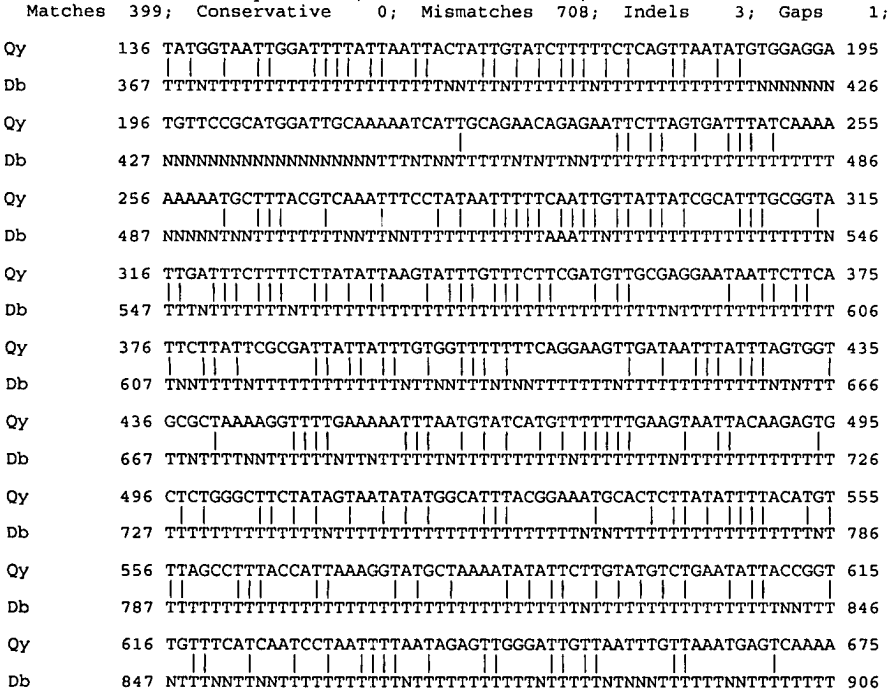


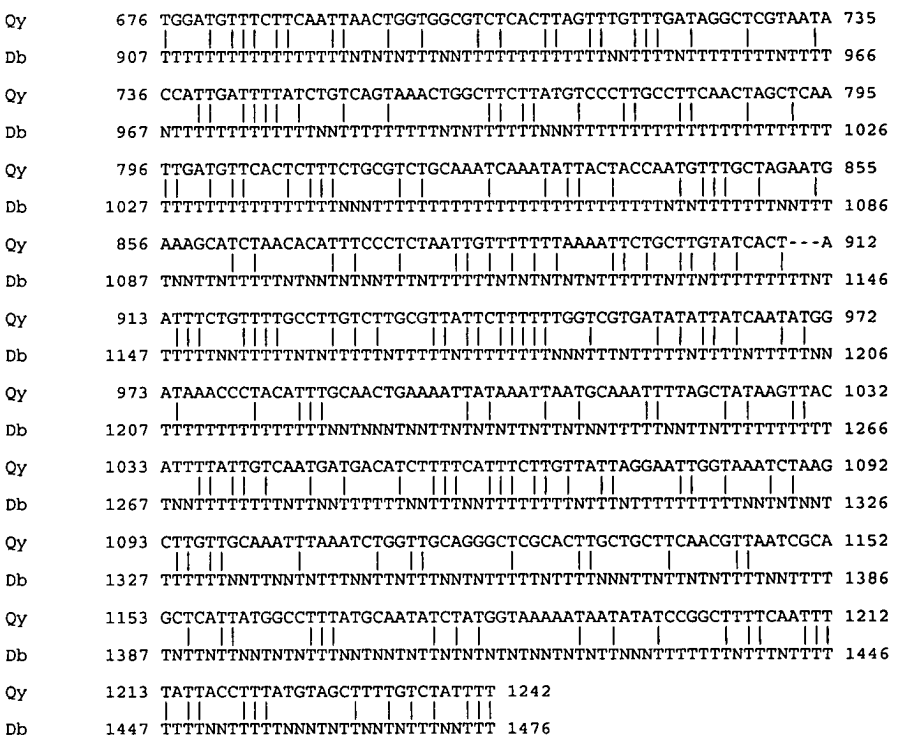


RESULT 7
AG435185
LOCUS AG435185 1843 bp DNA linear GSS 21-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-313C22.T7, genomic survey sequence.
ACCESSION AG435185
VERSION AG435185.1 GI:48078248
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823

```
REFERENCE   2  (bases 1 to 1843)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1843
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-313C22.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             8.2%;  Score 103.4;  DB 10;  Length 1843;
  Best Local Similarity   35.9%;  Pred. No. 5.9e-10;
  Matches  399;  Conservative    0;  Mismatches  708;  Indels    3;  Gaps    1;
```





```
RESULT 8
CL082569/c
LOCUS       CL082569               1632 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-167P5_Sp6.1 CH216 Xenopus tropicalis genomic clone
            CH216-167P5, genomic survey sequence.
ACCESSION   CL082569
VERSION     CL082569.1  GI:40538482
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1632)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 1062
            High quality sequence stop: 1114.
FEATURES             Location/Qualifiers
     source          1. .1632
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
```

```
                     /db_xref="taxon:8364"
                     /clone="CH216-167P5"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             8.1%;  Score 102.6;  DB 10;  Length 1632;
  Best Local Similarity   43.3%;  Pred. No. 8.5e-10;
  Matches  477;  Conservative    0;  Mismatches  622;  Indels    2;  Gaps    1;

Qy        144 TTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTGGAGGATGTTCCGC 203
Db       1547 TTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTATNTTTTNTTTTTTTT 1488

Qy        204 ATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGATTTATCAAAAAAAAATGC 263
Db       1487 ATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTC 1428

Qy        264 TTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTC 323
Db       1427 ATTTTTTTTTTTTTNTTTTTTTTTTTNATTNTTTTTTATTTAATTTTTTTTTTTTTATTTT 1368

Qy        324 TTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATT 383
Db       1367 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATNTNATTTTTNTTTTTTTTT 1308

Qy        384 CGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAA 443
Db       1307 ATTTTTTTTTTTATTATTTTTTTTNATTTTATTTTTTTTATTTTTTAATTTTTTTTTTTTT 1248

Qy        444 AGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGC 503
Db       1247 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTATATTTTTATTTNT 1188

Qy        504 TTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTT 563
Db       1187 NTTATTTTTTTTTTATTTTTTTTTTTTTTTTTATTTTTTTTTTTTTAATTATTTATTTTT 1128

Qy        564 TACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCAT 623
Db       1127 TTTTTATTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTT 1068

Qy        624 CAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTT 683
Db       1067 TTTTTTTATTTTTTTATTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTT 1008

Qy        684 TCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGAT 743
Db       1007 TTTTTTTTTTTTTTTTTTTTTTTAATATATTTTTTTTTTTTTTTTTTTAATATTTTTTTT 948

Qy        744 TTTATCTGTCAGTAAACTGGCT--TCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATG 801
Db        947 TTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTA 888

Qy        802 TTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCA 861
Db        887 TTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTATATTTATTTTT 828

Qy        862 TCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTT 921
Db        827 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTATTTTTTTTTTTTTTATTTTTTTTTT 768

Qy        922 TTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATAAACCCT 981
Db        767 TTTAAATTATTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTATTTTTTAT 708

Qy        982 ACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTTTATTG 1041
Db        707 TTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNATNTTTTTTT 648

Qy       1042 TCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATTGGTAAATCTAAGCTTGTTGCA 1101
Db        647 TTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTATTTTTTTTTTTT 588

Qy       1102 AATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCAACGTTAATCGCAGCTCATTAT 1161
Db        587 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTAATAAAAAATT 528

Qy       1162 GGCCTTTATGCAATATCTATGGTAAAAATAATATATCCGGCTTTTCAATTTTATTACCTT 1221
Db        527 TTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTNTTTTTT 468

Qy       1222 TATGTAGCTTTTGTCTATTTT 1242
Db        467 NTTTTTTTTTTTTTTTTTTT 447

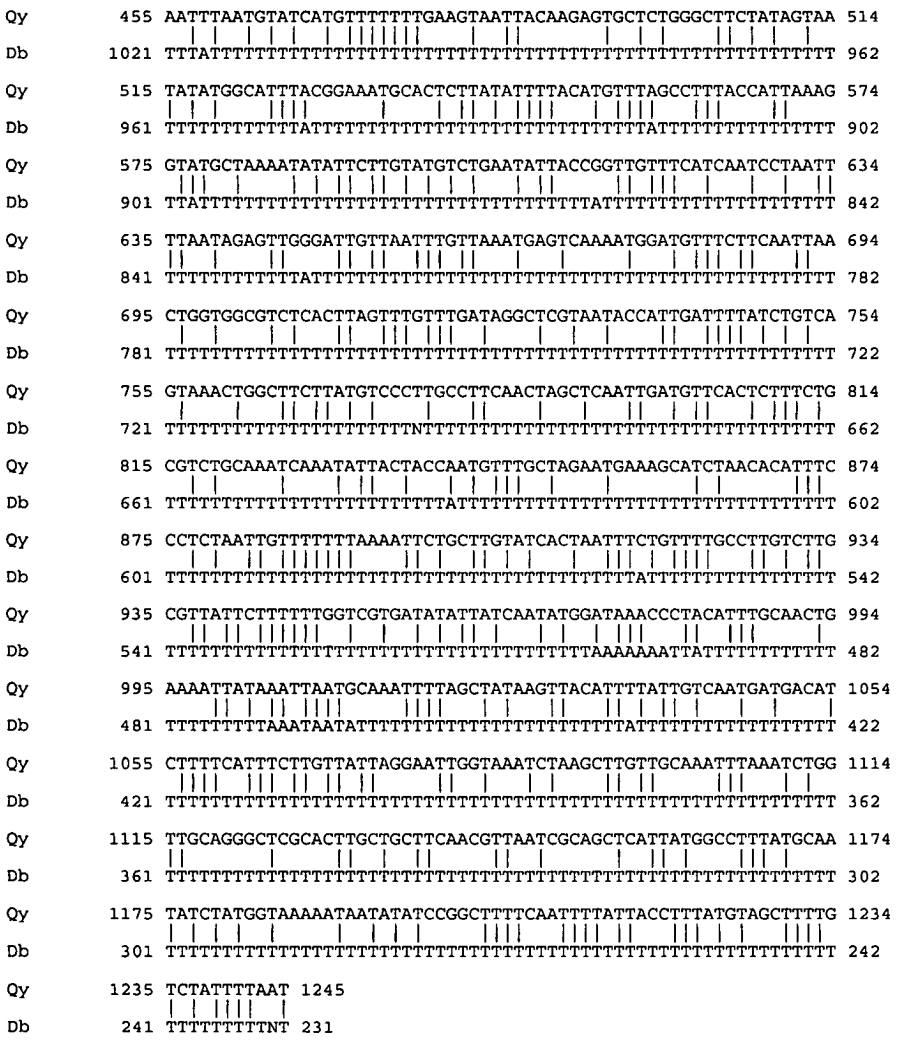

RESULT 9
CL016663/c
LOCUS       CL016663                1243 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-1A23_Sp6.3 CH216 Xenopus tropicalis genomic clone CH216-1A23,
            genomic survey sequence.
ACCESSION   CL016663
VERSION     CL016663.1  GI:40458476
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1243)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 702
            High quality sequence stop: 896.
FEATURES             Location/Qualifiers
     source          1. .1243
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-1A23"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             8.1%;  Score 102.4;  DB 10;  Length 1243;
  Best Local Similarity   44.0%;  Pred. No. 9.2e-10;
  Matches  427;  Conservative    0;  Mismatches  544;  Indels    0;  Gaps    0;

Qy        275 TTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATAT 334
Db       1201 TTTTTTTTTTTTTTTTTTAATTATTTATTATTTTTTTTTATTTTTTTTTTTTTTTATT 1142

Qy        335 TAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTA 394
Db       1141 TTATTTTTTTTTTTTTATTTTTTTTTAATTTTTATTTTTTTTTTTTTTTTTTTTTTTT 1082

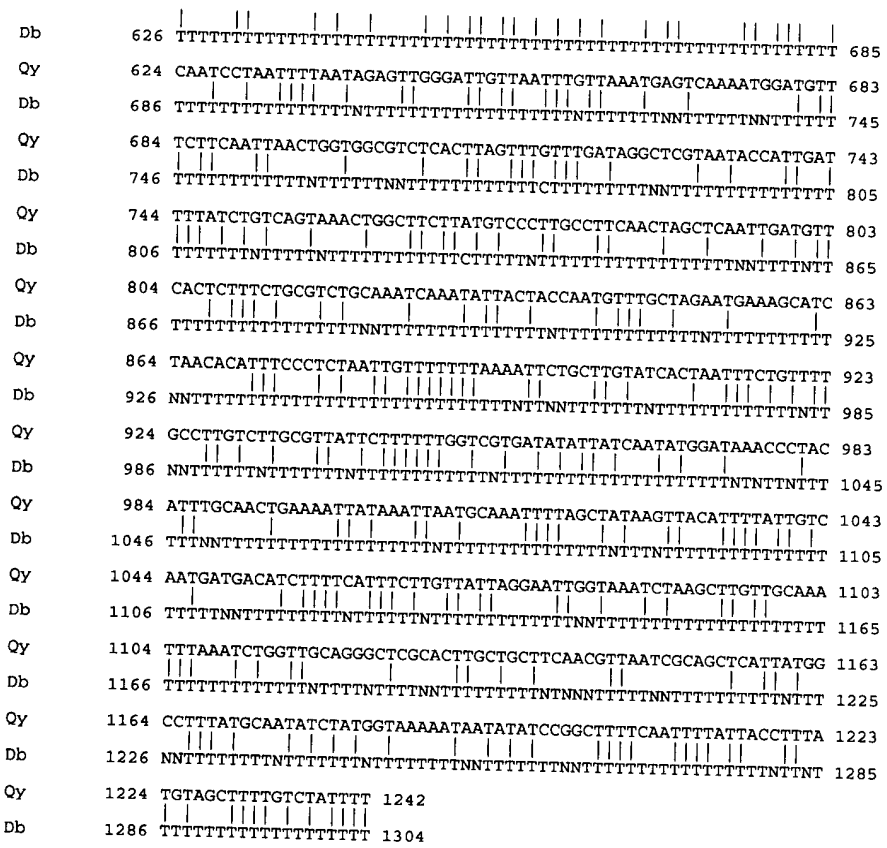
Qy        395 TTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAA 454
Db       1081 TTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTANTTTTATTTTTTATTT 1022
```



RESULT 10
AG386893
LOCUS AG386893 1519 bp DNA linear GSS 21-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-201E07.TJ, genomic survey sequence.
ACCESSION AG386893
VERSION AG386893.1 GI:47998098
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1519)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1519
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-201E07.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Query Match 8.1%; Score 102.4; DB 10; Length 1519;
Best Local Similarity 40.8%; Pred. No. 9.3e-10;
Matches 448; Conservative 0; Mismatches 651; Indels 0; Gaps 0;





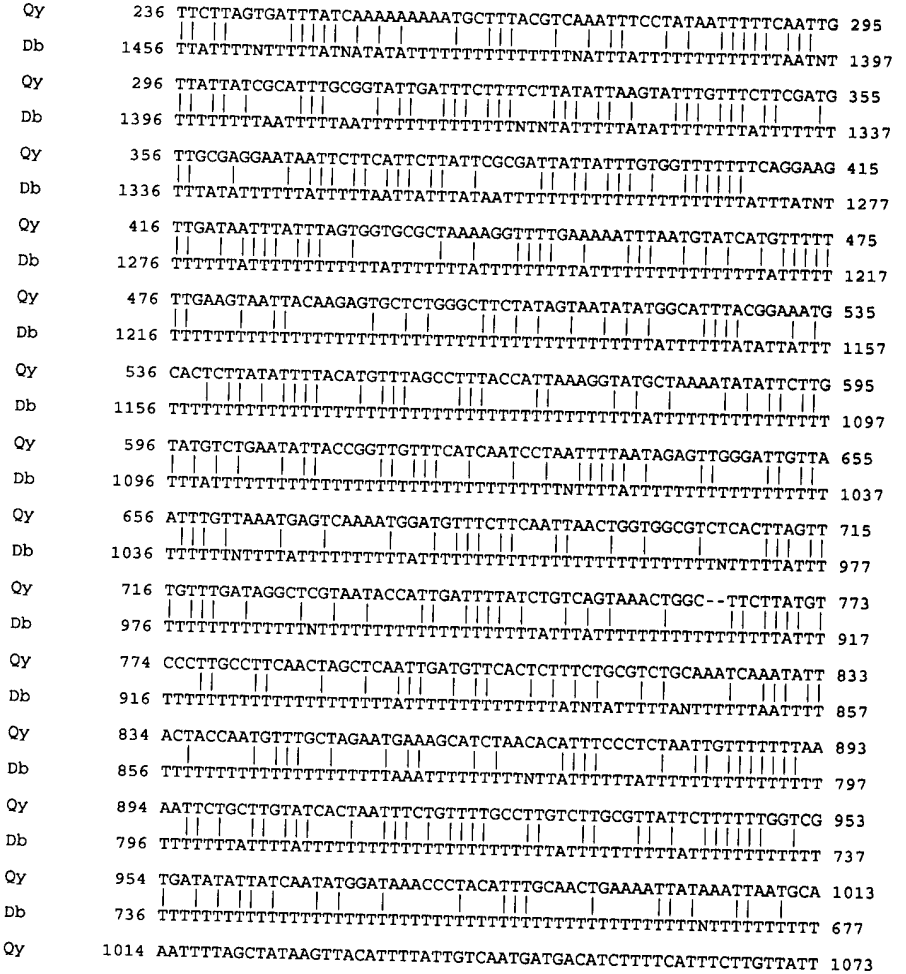
RESULT 11
CL078538/c
LOCUS CL078538 1536 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-151C11_Sp5.1 CH216 Xenopus tropicalis genomic clone CH216-151C11, genomic survey sequence.
ACCESSION CL078538
VERSION CL078538.1 GI:40534451
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1536)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp5 atctgccgtttcgatcct
Class: BAC ends

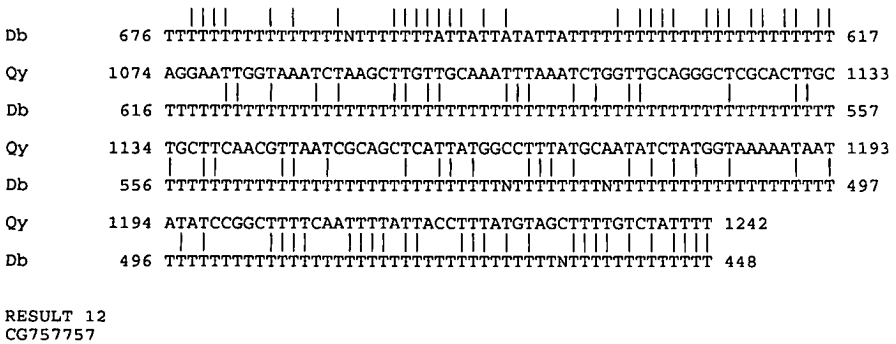
High quality sequence start: 1061
High quality sequence stop: 1145.

FEATURES Location/Qualifiers
source 1. .1536
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-151C11"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"

ORIGIN

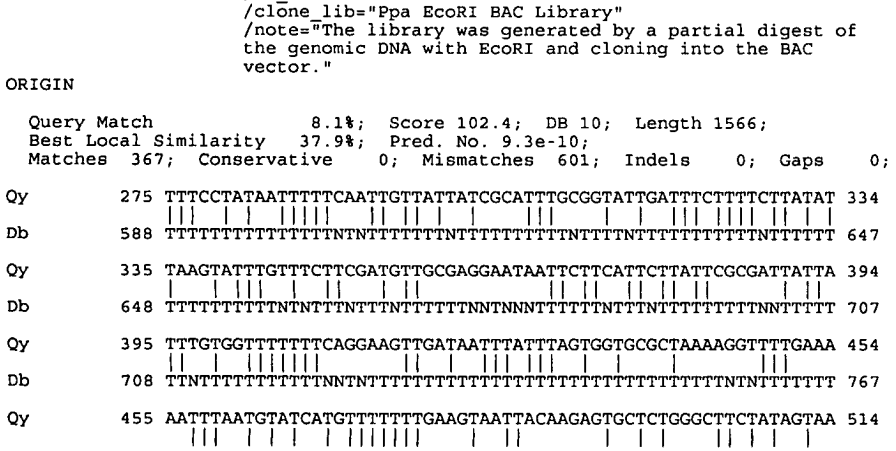
Query Match 8.1%; Score 102.4; DB 10; Length 1536;
Best Local Similarity 43.8%; Pred. No. 9.3e-10;
Matches 442; Conservative 0; Mismatches 565; Indels 2; Gaps 1;

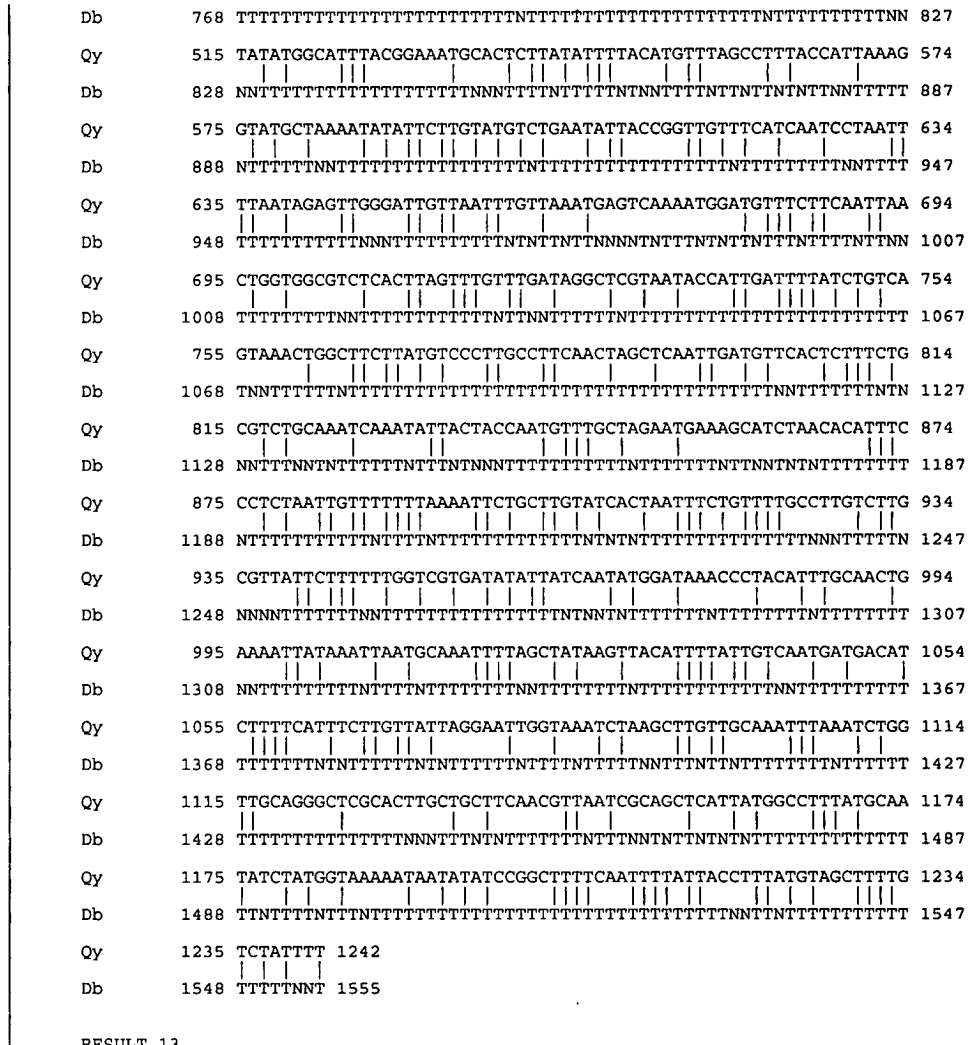




RESULT 12
CG757757
LOCUS CG757757 1566 bp DNA linear GSS 24-OCT-2003
DEFINITION P053-1-D07.zc Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG757757
VERSION CG757757.1 GI:37986636
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1566)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1566
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."
ORIGIN

Query Match 8.1%; Score 102.4; DB 10; Length 1566;
Best Local Similarity 37.9%; Pred. No. 9.3e-10;
Matches 367; Conservative 0; Mismatches 601; Indels 0; Gaps 0;





RESULT 13
CL082000
LOCUS CL082000 1981 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-165P18_RM4.1 CH216 Xenopus tropicalis genomic clone CH216-165P18, genomic survey sequence.
ACCESSION CL082000
VERSION CL082000.1 GI:40537913
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1981)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)

```
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000    Std Error: 0.00
            Seq primer: RM4 ctcaagggcatcggtcgagc
            Class: BAC ends
            High quality sequence start: 265
            High quality sequence stop: 497.
FEATURES             Location/Qualifiers
     source          1. .1981
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-165P18"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN
```

Query Match 8.1%; Score 102.2; DB 10; Length 1981;
Best Local Similarity 45.5%; Pred. No. 1e-09;
Matches 495; Conservative 0; Mismatches 578; Indels 16; Gaps 3;

```
Qy        144 TTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTGGAGGATGTTCCGC 203
Db        536 TTTTATTTTTTTTTTTTATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 595

Qy        204 ATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGATTTATCAAAAAAAAATGC 263
Db        596 TTTTTATTTTTTTTTTATTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 655

Qy        264 TTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTC 323
Db        656 TTTTTTTTAGTTTTATTATTTTTTTTTATTTTTTTTTATTATTTTTTTTTTTTTTTTTA 715

Qy        324 TTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATT 383
Db        716 TTATTTTTTTTTATTTTTTTTTTGTTTTTTTATTTTTTTTTTTATTTTTTTTTATTTTTTT 775

Qy        384 CGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAA 443
Db        776 TTAATTTTATTTTTTTTATTGTTTTATTTTGTTTTTTTTATTTTATTTTTTTTTTTTTTT 835

Qy        444 AGGTTTTGAAAAATTTAATGTATCATGTTTT-TTTGAAGTAATTACAAGAGTGCTCTGGG 502
Db        836 TTTTTTTTTTTTTTTTTATTTTTTTTATATTCTTTTTTTTTATTTGTTATTTATTTTTTT 895

Qy        503 CTTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCT 562
Db        896 ATTTTTTTTTTATTATTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 955

Qy        563 TTACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCA 622
Db        956 TTTTTTATTTATATTTATTTTATTATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1015

Qy        623 TCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGT 682
Db       1016 TTTTTATTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1075

Qy        683 TTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGA 742
Db       1076 TTTTTTTTTTTA----------TATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1124

Qy        743 TTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATGT 802
Db       1125 TTTTATATTTTTTTTATTTTATTTTTTATTTTATTATTTGTTATTTATTTTTTTTTATTT 1184

Qy        803 TCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCAT 862
```

```
Db       1185 TTAATTATTTATTTTTTATATTTATTTTATTATTTTACTTGTATTTTATTTTTTTTTTTT 1244

Qy        863 CTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTTT 922
Db       1245 TTTTTATTTTTTTTTATTTTTTATTGTTTTATTTTTTTTTTTATTTTCTTTTTTATTTTT 1304

Qy        923 TGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATAAACCCTA 982
Db       1305 ATTTTTTTTTTTATTTTTTTATTTTTATTTCTTTTATTTTTATTTTTTTTTTATTTTCTA 1364

Qy        983 CATTTGCAACTGAAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTTTATTGT 1042
Db       1365 TTTATTTATTT----ATTTTATTATATATTTTTTTTTTTTATATTTTATTTTGTTTATTAA 1420

Qy       1043 CAATGATGACATCTTTTCATTTCTTGTTATTAGGAATTGGTAAATCTAAGCTTGTTGCAA 1102
Db       1421 TTATTATATTTTTTTATTTTTTTTTTTTATTTTTATATATTTTTTTTTATGTTTTTTATTT 1480

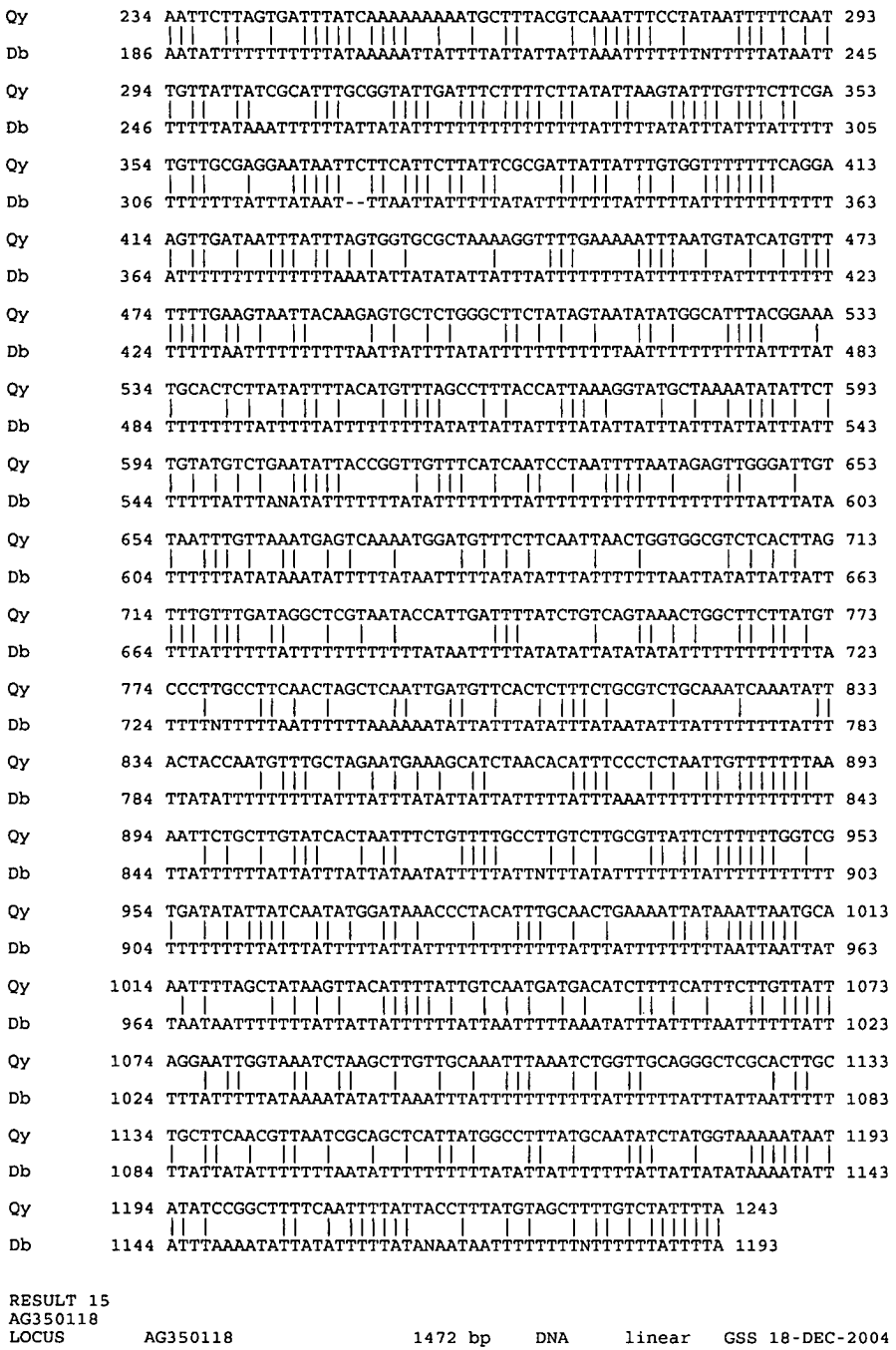
Qy       1103 ATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCAACGTTAATCGCAGCTCATTATG 1162
Db       1481 TTTTTTTTTTTATTGATTGTTTATTTTTTTTATTTTTTTATTTATTTTTTTTTTTTTATT 1540

Qy       1163 GCCTTTATGCAATATCTATGGTAAAAATAATATATCCGGCTTTTCAATTTTATTACCTTT 1222
Db       1541 TTTTTTATATATTATTTTTTTTTTATTATTTTTTATATTTTGTTTTTTTTTTTTTTTTT 1600

Qy       1223 ATGTAGCTT 1231
Db       1601 TTATATATT 1609
```

```
RESULT 14
CG749499
LOCUS       CG749499                1348 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P043-4-A06.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG749499
VERSION     CG749499.1  GI:37970425
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1348)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1348
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN
```

Query Match 8.1%; Score 102; DB 10; Length 1348;
Best Local Similarity 44.3%; Pred. No. 1.1e-09;
Matches 447; Conservative 0; Mismatches 561; Indels 2; Gaps 1;

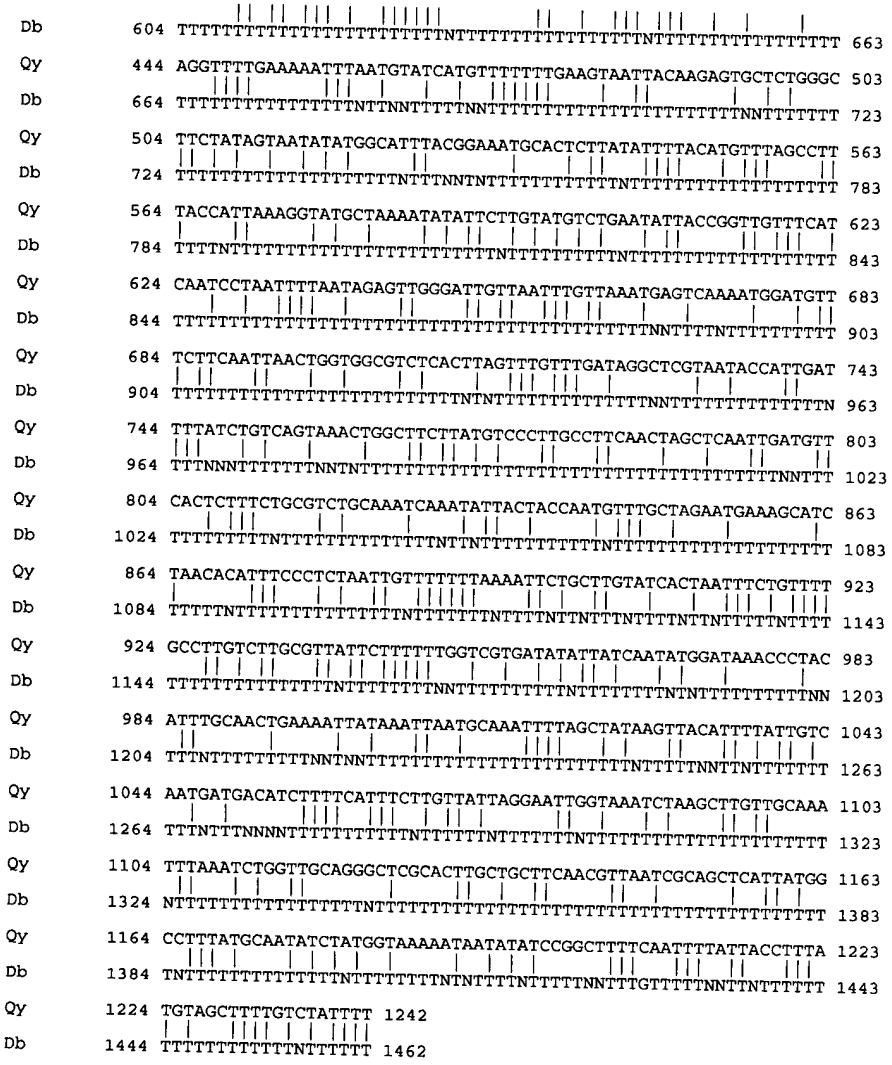


RESULT 15
AG350118
LOCUS AG350118 1472 bp DNA linear GSS 18-DEC-2004

DEFINITION Mus musculus molossinus DNA, clone:MSMg01-146K03.TJ, genomic survey sequence.
ACCESSION AG350118
VERSION AG350118.1 GI:47923428
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1472)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1472
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-146K03.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Query Match 8.0%; Score 101.6; DB 10; Length 1472;
Best Local Similarity 38.7%; Pred. No. 1.3e-09;
Matches 425; Conservative 0; Mismatches 674; Indels 0; Gaps 0;





Search completed: May 11, 2006, 05:00:50
Job time : 4058.75 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 20:34:42 ; Search time 176.589 Seconds
(without alignments)
12743.667 Million cell updates/sec

Title: US-09-423-093B-1_COPY_8646_9911
Perfect score: 1266
Sequence: 1 atggtattaacagtgaaaaa.....gagcgaaaaatgtctattga 1266

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	77.4	6.1	1039	3	US-09-902-540-1280	Sequence 1280, Ap
c	2	71.2	5.6	612	3	US-09-902-540-1357	Sequence 1357, Ap
	3	67.6	5.3	50000	3	US-09-662-254B-24	Sequence 24, Appl
	4	66.6	5.3	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	5	64.6	5.1	187169	3	US-09-949-016-12776	Sequence 12776, A
c	6	64.6	5.1	191569	3	US-09-949-016-15940	Sequence 15940, A
c	7	64	5.1	32392	3	US-09-662-254B-27	Sequence 27, Appl
	8	63.8	5.0	19438	3	US-09-949-016-12699	Sequence 12699, A
c	9	63.4	5.0	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	10	58.8	4.6	1830	3	US-09-662-254B-67	Sequence 67, Appl
	11	58.8	4.6	32392	3	US-09-662-254B-27	Sequence 27, Appl
c	12	57.6	4.5	640681	3	US-09-790-988-1	Sequence 1, Appli
	13	55.6	4.4	7218	2	US-08-232-463-14	Sequence 14, Appl
	14	55.4	4.4	615	3	US-08-998-416-186	Sequence 186, App
	15	54.4	4.3	147382	3	US-09-949-016-14624	Sequence 14624, A
	16	54.2	4.3	18773	3	US-09-949-016-14164	Sequence 14164, A
	17	53	4.2	636	3	US-08-998-416-1137	Sequence 1137, Ap
	18	52.8	4.2	825	3	US-09-134-001C-207	Sequence 207, App
c	19	52.8	4.2	4069	3	US-09-710-279-3976	Sequence 3976, Ap
c	20	52.6	4.2	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	21	52.6	4.2	9636	2	US-08-323-170B-1	Sequence 1, Appli
c	22	52.6	4.2	9636	3	US-08-954-441-1	Sequence 1, Appli
	23	52.4	4.1	50000	3	US-09-662-254B-23	Sequence 23, Appl
	24	52.4	4.1	50000	3	US-09-662-254B-25	Sequence 25, Appl
	25	51.6	4.1	837	3	US-08-998-416-288	Sequence 288, App
c	26	51.4	4.1	50000	3	US-09-662-254B-24	Sequence 24, Appl
c	27	50.2	4.0	240	2	US-08-628-417-6	Sequence 6, Appli
c	28	50.2	4.0	614	3	US-09-902-540-1318	Sequence 1318, Ap
c	29	49.6	3.9	714	3	US-09-662-254B-51	Sequence 51, Appl
	30	49.4	3.9	396	3	US-09-640-173-53	Sequence 53, Appl
	31	49.4	3.9	396	3	US-09-713-550-53	Sequence 53, Appl
	32	49.4	3.9	396	3	US-09-825-294-53	Sequence 53, Appl
	33	49.4	3.9	396	3	US-09-970-966-53	Sequence 53, Appl
c	34	49.4	3.9	700	3	US-09-735-271-1038	Sequence 1038, Ap
c	35	49.4	3.9	1445	3	US-09-830-230A-548	Sequence 548, App
c	36	49.4	3.9	1493	3	US-09-830-230A-547	Sequence 547, App
	37	49.4	3.9	1905	3	US-09-270-767-10264	Sequence 10264, A
	38	49.4	3.9	640681	3	US-09-790-988-1	Sequence 1, Appli
	39	49.2	3.9	658	3	US-08-998-416-595	Sequence 595, App
c	40	49.2	3.9	10640	3	US-09-417-485D-5	Sequence 5, Appli
c	41	49.2	3.9	50000	3	US-09-662-254B-26	Sequence 26, Appl
	42	49	3.9	5340	3	US-09-627-122-21	Sequence 21, Appl
	43	48.8	3.9	1431	3	US-09-316-083-2	Sequence 2, Appli
	44	48.8	3.9	1431	3	US-09-933-700-2	Sequence 2, Appli
c	45	48.8	3.9	3275	3	US-09-370-838-151	Sequence 151, App

ALIGNMENTS

RESULT 1
US-09-902-540-1280/c
; Sequence 1280, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1280
; LENGTH: 1039
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1039)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1280

Query Match 6.1%; Score 77.4; DB 3; Length 1039;
Best Local Similarity 45.6%; Pred. No. 1.2e-07;
Matches 425; Conservative 0; Mismatches 498; Indels 10; Gaps 4;

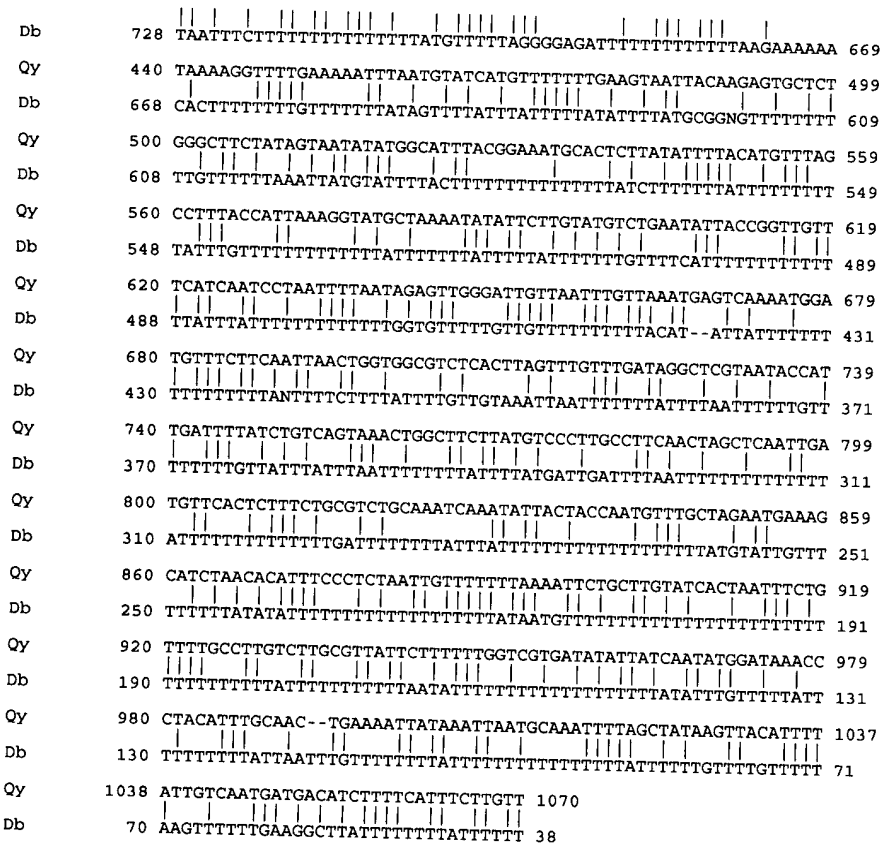
```
Qy        144 TTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTGGAGGATGTTCCGC 203
              ||   |||| |||      |  | | | ||||| |  | || | |||       | ||   |
Db        966 TTTTTTTTTTTTATGGTTTTATTTTTTTTTTTTTTGGATATTTTGTAATTTTTTTTTTC 907

Qy        204 ATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGATTTATCAAAAAAAAATGC 263
              |  |||      | |  ||             ||| ||  | | ||||          |
Db        906 TTTTATTTTTTTATTTTTTATTTGGTTTTTTATTTTTTTTTAAGTTATTTTTGTTTTTTAT 847

Qy        264 TTTACGTCAAATTTCCTATAATT----TTTCAATTGTTATTATCGCATTTGCGGTATTGA 319
              ||| | |||  |||    |  ||     |||  || || |||      |||     | ||
Db        846 TTTCCCTCATTTTTTTAGTGTTTTTAATTTTTTTTTTTTTTTAGAAGGTTTATATTTTTAT 787

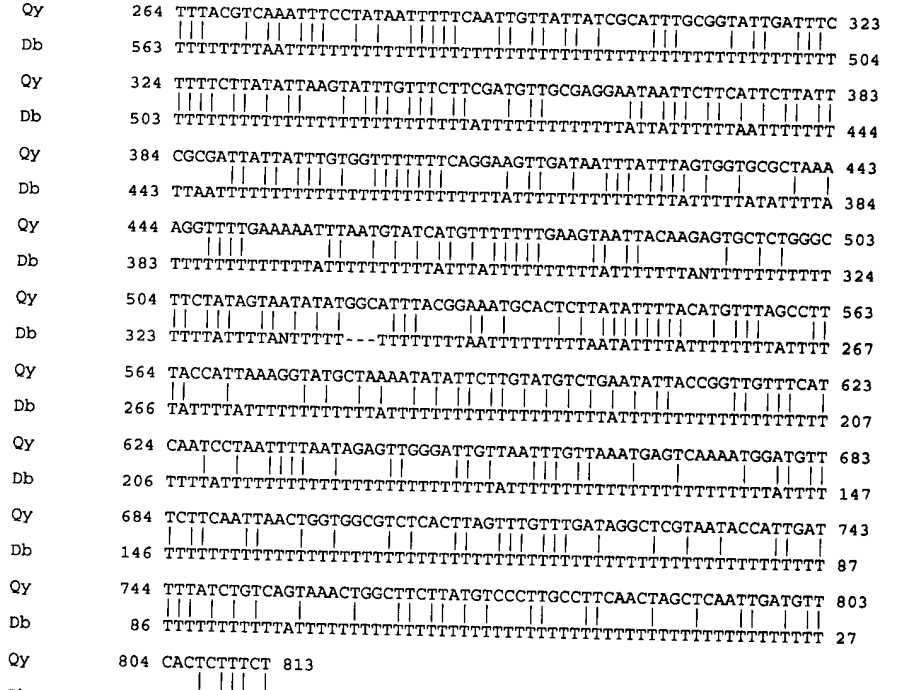
Qy        320 TTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCT 379
              || |||| || | || || |||| ||  |    ||  | |  || ||   |  | || |
Db        786 ATTTTTTTTTTGTTTTTAGGATTTTTTGTTATATTGGGGTGTTGATTAGAGC--CTTTGT 729

Qy        380 TATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGC 439
```



RESULT 2
US-09-902-540-1357/c
; Sequence 1357, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1357
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(612)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1357

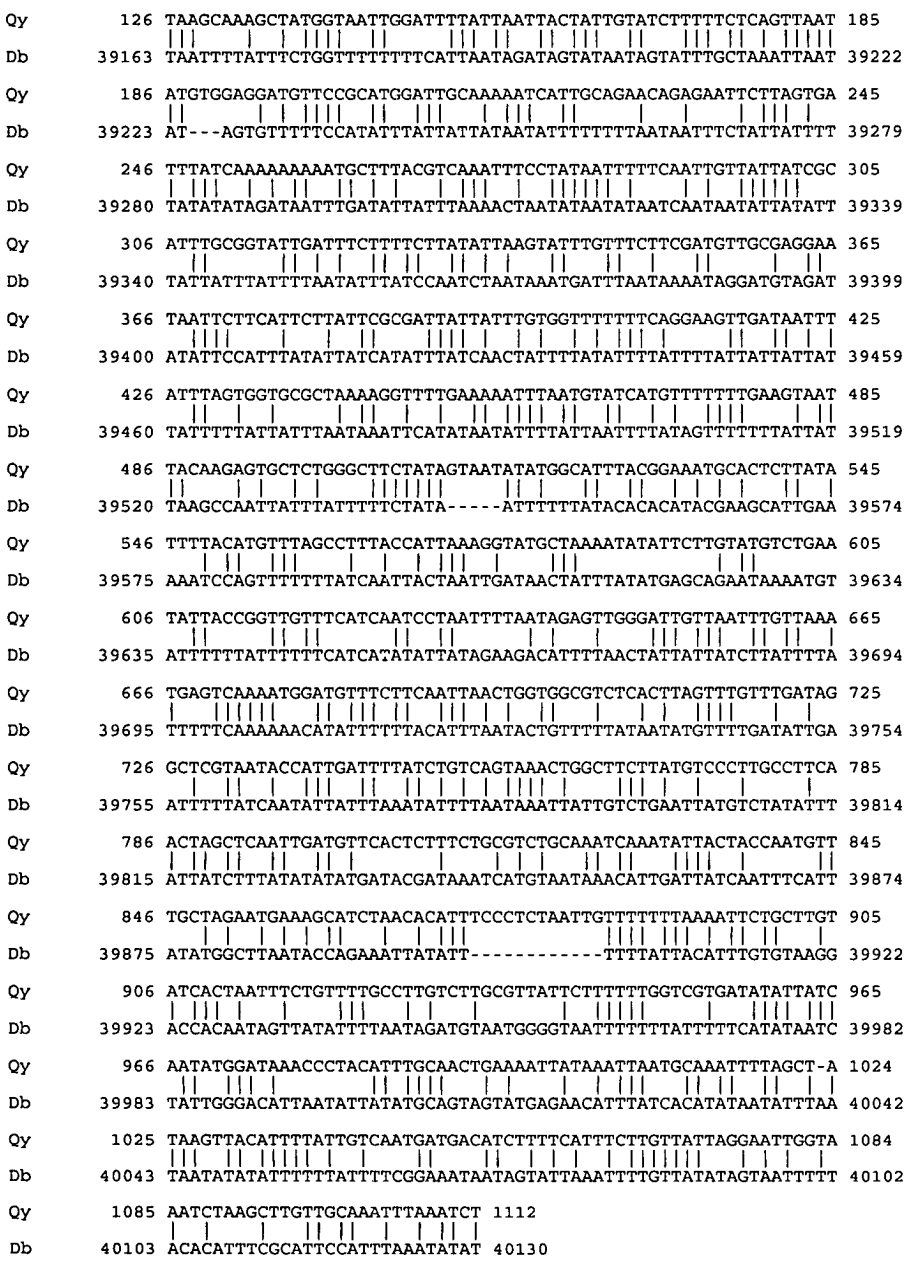
Query Match 5.6%; Score 71.2; DB 3; Length 612;
Best Local Similarity 46.7%; Pred. No. 2.4e-06;
Matches 257; Conservative 0; Mismatches 290; Indels 3; Gaps 1;



RESULT 3
US-09-662-254B-24
; Sequence 24, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-24

Query Match 5.3%; Score 67.6; DB 3; Length 50000;
Best Local Similarity 44.8%; Pred. No. 3.5e-05;

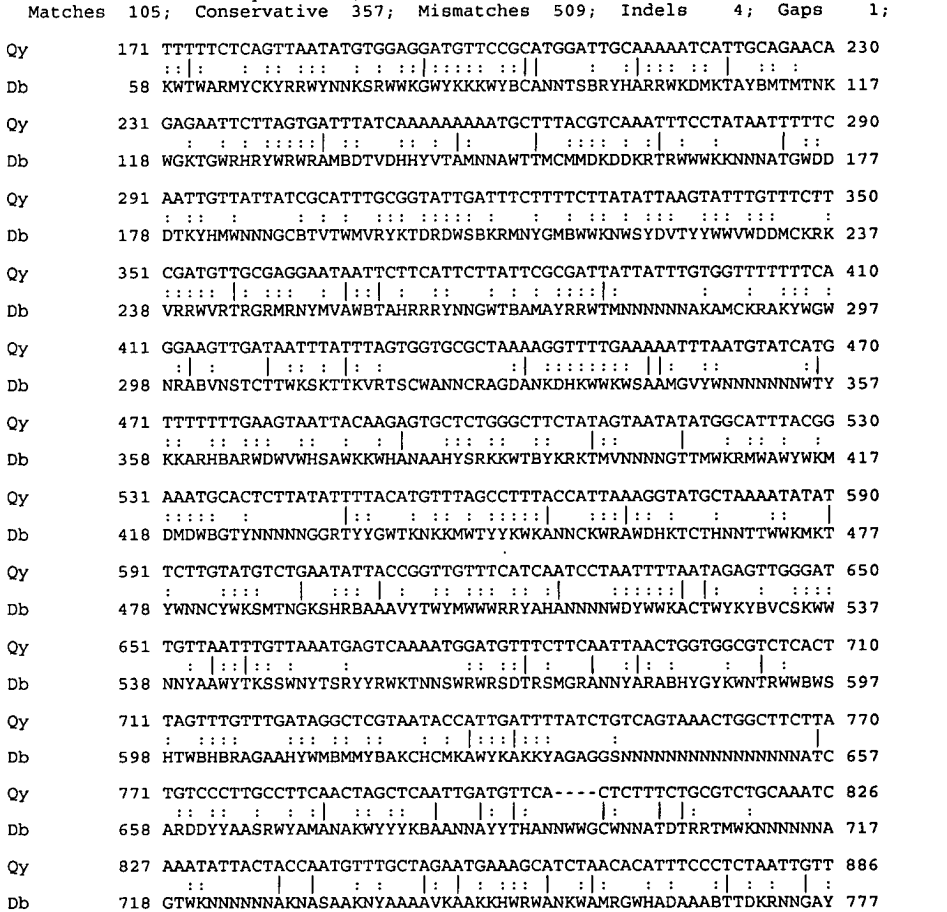
Matches 443; Conservative 0; Mismatches 524; Indels 21; Gaps 4;

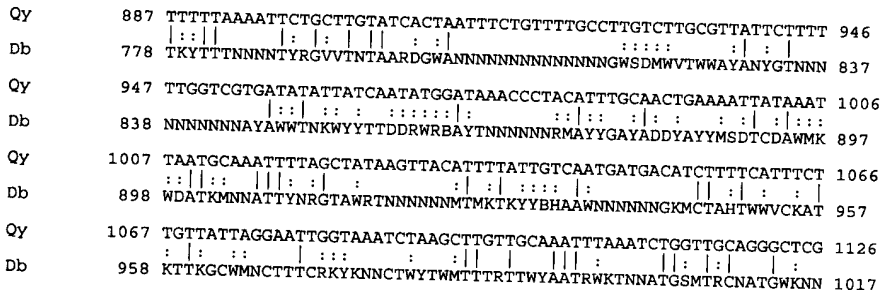


RESULT 4
US-09-806-708B-22

; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 5.3%; Score 66.6; DB 3; Length 1141;
Best Local Similarity 10.8%; Pred. No. 2.7e-05;
Matches 105; Conservative 357; Mismatches 509; Indels 4; Gaps 1;





RESULT 5
US-09-949-016-12776/c
; Sequence 12776, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12776
; LENGTH: 187169
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(187169)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12776

Query Match 5.1%; Score 64.6; DB 3; Length 187169;
Best Local Similarity 43.9%; Pred. No. 0.0002;
Matches 379; Conservative 0; Mismatches 474; Indels 10; Gaps 2;

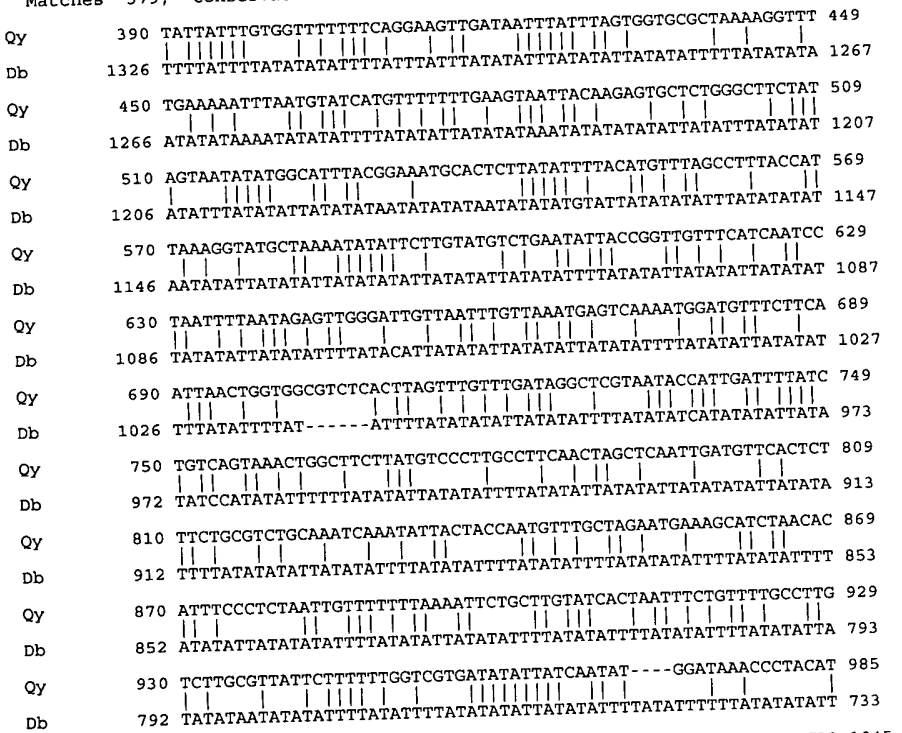




RESULT 6
US-09-949-016-15940/c
; Sequence 15940, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15940
; LENGTH: 191569
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(191569)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15940

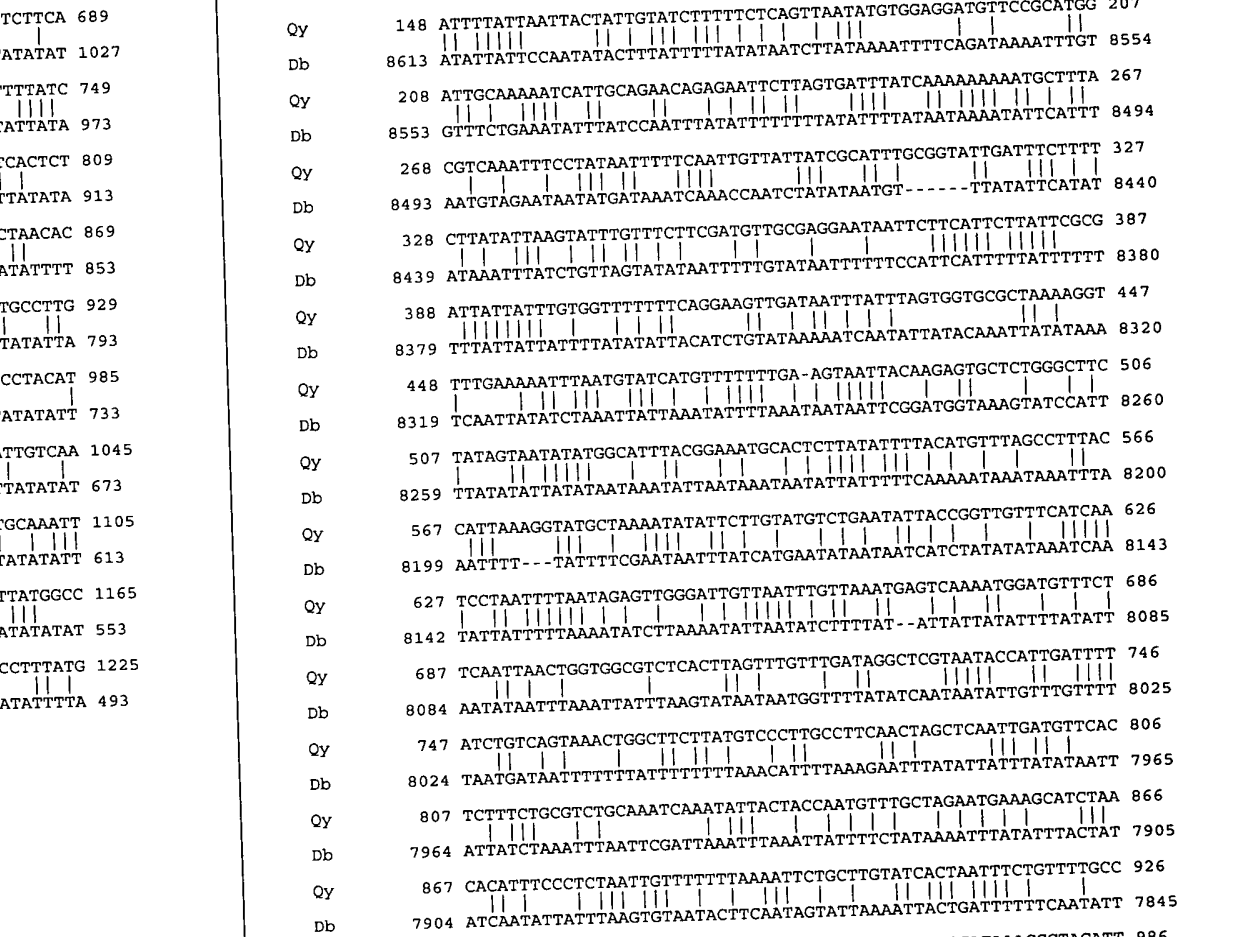
Query Match 5.1%; Score 64.6; DB 3; Length 191569;

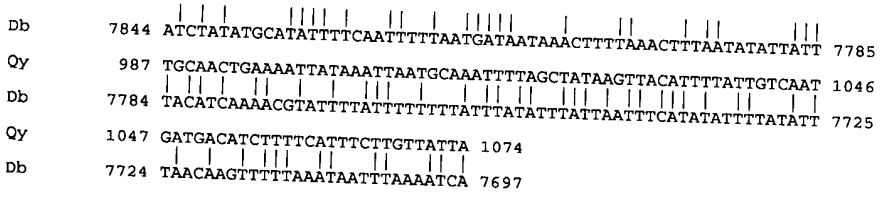
Best Local Similarity 43.9%; Pred. No. 0.0002;
Matches 379; Conservative 0; Mismatches 474; Indels 10; Gaps 2;



RESULT 7
US-09-662-254B-27/c
; Sequence 27, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 32392
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-27

Query Match 5.1%; Score 64; DB 3; Length 32392;
Best Local Similarity 44.8%; Pred. No. 0.00019;
Matches 416; Conservative 0; Mismatches 500; Indels 12; Gaps 4;





```
Db       7844 ATCTATATGCATATTTTCAATTTTTAATGATAATAAACTTTTAAACTTTAATATATTATT 7785

Qy        987 TGCAACTGAAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTTTATTGTCAAT 1046

Db       7784 TACATCAAAACGTATTTTATTTTTTTTATTTATATTTATTAATTTCATATATTTTATATT 7725

Qy       1047 GATGACATCTTTTCATTTCTTGTTATTA 1074

Db       7724 TAACAAGTTTTTAAATAATTTAAAATCA 7697
```

RESULT 8
US-09-949-016-12699
; Sequence 12699, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12699
; LENGTH: 19438
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12699

Query Match 5.0%; Score 63.8; DB 3; Length 19438;
Best Local Similarity 46.0%; Pred. No. 0.00019;
Matches 215; Conservative 0; Mismatches 252; Indels 0; Gaps 0;



```
Qy        143 ATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTGGAGGATGTTCCG 202

Db      18490 ATTTATTATATTAAATATTTATTATAATATTTAATTTATTTATATATAATTTAATTTATA 18549

Qy        203 CATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGATTTATCAAAAAAAAATG 262

Db      18550 TATAATTTAATTACATAATTATTTTTATAATTATATATAATTATTTTTAATTAATTAATT 18609

Qy        263 CTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTT 322

Db      18610 ATATATAATTAATTATATATAATTATATATAATTATATATATAATTAAATATAATTATTT 18669

Qy        323 CTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTAT 382

Db      18670 ATATAATTATATGTAATAATATAATTATATATAATTATATATATAATTAAATATAATTAT 18729

Qy        383 TCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAA 442

Db      18730 TATATATACTTATATATAATTAAATATATAAAAATATATAATTAAATATATATAATTTAA 18789

Qy        443 AAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGG 502

Db      18790 TTATATAATTATTATATTATATTAAATATTTATTATATTTCATTTATTTATATATATAAT 18849

Qy        503 CTTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCT 562

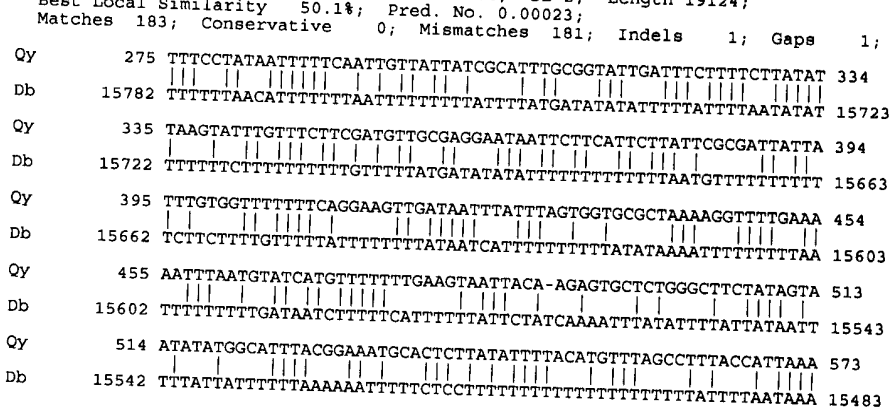
Db      18850 TTAATTTATACATATAATTTAATTAAATTATATATAATAAATATATTAAATATTTATTAT 18909

Qy        563 TTACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATT 609

Db      18910 ATTTATTTTAATATATTAATTATATTAATATTTAATAATTTAATAAT 18956
```

RESULT 9
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellems, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

Query Match 5.0%; Score 63.4; DB 2; Length 19124;
Best Local Similarity 50.1%; Pred. No. 0.00023;
Matches 183; Conservative 0; Mismatches 181; Indels 1; Gaps 1;



```
Qy        275 TTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATAT 334

Db      15782 TTTTTTAACATTTTTTTAATTTTTTTTTATTTTATGATATATATTTTTATTTTAATATAT 15723

Qy        335 TAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTA 394

Db      15722 TTTTTTCTTTTTTTTTTTGTTTTTATGATATATATTTTTTTTTTTTTAATGTTTTTTTTTT 15663

Qy        395 TTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAA 454

Db      15662 TCTTCTTTTGTTTTTATTTTTTTTATAATCATTTTTTTTTTATATAAAATTTTTTTTTAA 15603

Qy        455 AATTTAATGTATCATGTTTTTTTGAAGTAATTACA-AGAGTGCTCTGGGCTTCTATAGTA 513

Db      15602 TTTTTTTTTGATAATCTTTTTCATTTTTTATTCTATCAAAATTTATATTTTATTATAATT 15543

Qy        514 ATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTTTACCATTAAA 573

Db      15542 TTTATTATTTTTTAAAAAATTTTTCTCCTTTTTTTTTTTTTTTTTTTTTTATTTTAATAAA 15483
```

```
Qy        574 GGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCATCAATCCTAAT 633
Db      15482 TTTTTTTTTATATTTCATTTTTCTTTTTTTCATTTTAATAAATTTGTTTTTATATTTCTT 15423

Qy        634 TTTAA 638
Db      15422 TTTTA 15418
```

RESULT 10
US-09-662-254B-67/c
; Sequence 67, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1)..(1830)
; OTHER INFORMATION:
US-09-662-254B-67

Query Match 4.6%; Score 58.8; DB 3; Length 1830;
Best Local Similarity 46.4%; Pred. No. 0.0015;
Matches 271; Conservative 0; Mismatches 302; Indels 11; Gaps 2;

```
Qy        131 AAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTG 190
Db       1483 AATTATTTTCTACATTAATATTATTTAATAAATTAAAATAAGGTTTATATTTTACATATT 1424

Qy        191 GAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGATTTAT 250
Db       1423 CTATAGCATCGTCAATATTTAAAAATATATTAGAAAAAAAGTAAATTATAAAATATTTAT 1364

Qy        251 CAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTG 310
Db       1363 AATAAAAATATATTTTAGGATAATTATAATTATTATTAAGCACACATATATTATTATCTG 1304

Qy        311 CGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATT 370
Db       1303 ATTTATATATATTTTTTGTATTTTTTATAATTATTTCATTATTTTTTATTTCAAATTTAT 1244

Qy        371 CTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTA 430
Db       1243 CTTCAATAACATTCATAAAAAATATCATATTATTAACATTAAAATTTAATATTTCTATA- 1185

Qy        431 GTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAA 490
Db       1184 --------TTATTAATATTTGTTATTTTATTTTTATTATTTGTGTTTATTTTACTATTAA 1133

Qy        491 GAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTA 550
Db       1132 ATAATGTTTCTGCTTTAAAAACAGGAATATTATTAATATCAAATTCATTATTTAAATATA 1073

Qy        551 CATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTA 610
Db       1072 GTTGATTAGGTATTTGTAAAAAAAATTTATTTATATCTGTAGATGAATTTATAATATAAT 1013

Qy        611 CCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGT 670
Db       1012 CAATATATTTATT--ATAATAAGAATAATATTTATTTATATTATTACTTTTTATATTATT 955

Qy        671 CAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGT 714
Db        954 ATAAATATTTATAATATCAAAATATTCATTTAATATTACATAAT 911
```

RESULT 11
US-09-662-254B-27
; Sequence 27, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 32392
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-27

Query Match 4.6%; Score 58.8; DB 3; Length 32392;
Best Local Similarity 46.4%; Pred. No. 0.0026;
Matches 271; Conservative 0; Mismatches 302; Indels 11; Gaps 2;

```
Qy        131 AAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTG 190
Db       9775 AATTATTTTCTACATTAATATTATTTAATAAATTAAAATAAGGTTTATATTTTACATATT 9834

Qy        191 GAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGATTTAT 250
Db       9835 CTATAGCATCGTCAATATTTAAAAATATATTAGAAAAAAAGTAAATTATAAAATATTTAT 9894

Qy        251 CAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTG 310
Db       9895 AATAAAAATATATTTTAGGATAATTATAATTATTATTAAGCACACATATATTATTATCTG 9954

Qy        311 CGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATT 370
Db       9955 ATTTATATATATTTTTTGTATTTTTTATAATTATTTCATTATTTTTTATTTCAAATTTAT 10014

Qy        371 CTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTA 430
Db      10015 CTTCAATAACATTCATAAAAAATATCATATTATTAACATTAAAATTTAATATTTCTATA- 10073

Qy        431 GTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAA 490
Db      10074 --------TTATTAATATTTGTTATTTTATTTTTATTATTTGTGTTTATTTTACTATTAA 10125

Qy        491 GAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTA 550
Db      10126 ATAATGTTTCTGCTTTAAAAACAGGAATATTATTAATATCAAATTCATTATTTAAATATA 10185

Qy        551 CATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTA 610
Db      10186 GTTGATTAGGTATTTGTAAAAAAAATTTATTTATATCTGTAGATGAATTTATAATATAAT 10245

Qy        611 CCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGT 670
```

```
Db      10246 CAATATATTTATT--ATAATAAGAATAATATTTATTTATATTATTACTTTTTATATTATT 10303

Qy        671 CAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGT 714

Db      10304 ATAAATATTTATAATATCAAAATATTCATTTAATATTACATAAT 10347
```

RESULT 12
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 4.5%; Score 57.6; DB 3; Length 640681;
Best Local Similarity 42.8%; Pred. No. 0.0083;
Matches 348; Conservative 0; Mismatches 464; Indels 2; Gaps 1;

```
Qy        297 TATTATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGT 356

Db     526436 TGTTTTCATATTTTATTATTGATTTTTCTCGTAAATCTTTGAGAATAGAATTAAAAATTA 526377

Qy        357 TGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGT 416

Db     526376 TTTCAGTATTATGTCTTGATAAATATTCAAGAATAACATTTTTTTCTTTTTCGGAGAAAT 526317

Qy        417 TGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTT 476

Db     526316 TTGTATTGTAAACTTTTTCGAGTGATATAATTATAAAATTTTTATTTTTATCGTTATATA 526257

Qy        477 TGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGC 536

Db     526256 AAGTATATATTTTCTTACCTTTTTGTGGATGAGGTAAAGAAAAAACTATCGATGTAATAG 526197

Qy        537 ACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGT 596

Db     526196 CACTTAGATCGTAACGTGATATAATTTCAGGATCAGTGAAATAAAGATTTGATTGTTTAA 526137

Qy        597 ATGTCTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAA 656

Db     526136 ATAATTTTTTTCTTCCTTGTGTTAAATCATAAATAATTTCTTCTGATTTTTTTTTGGT-- 526079

Qy        657 TTTGTTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTT 716

Db     526078 TTCTTTTATTGCTTTTATAAACCTTAATTTTTTTATAATATTTTTTTTAACATTTTCGAA 526019

Qy        717 GTTTGATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCC 776

Db     526018 ATGTTGAATTTCTTTATTTTTAAATTTTTTAATTTTAATTAAAAAAGATTGATTTTTTTT 525959

Qy        777 TTGCCTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACT 836

Db     525958 TAAAACAATAAAATGAAACTGCGGTTTGACTGTTGTATCTTTTTGAAATAGTTTTTTGTT 525899

Qy        837 ACCAATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAAT 896

Db     525898 AAAAATGAATTGTTTTAATATAGGAATATTTAATACTTTAGGTATTGATTTTTTATCAAA 525839
```

```
Qy        897 TCTGCTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGA 956

Db     525838 CCAATCAGTTTCTTGAATTAATATGTTATTCTCTTTAAGAATCCTCTCAATTTTTCCTGG 525779

Qy        957 TATATTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAAT 1016

Db     525778 ATCGTTTTTAAGATTATTAGATATTTTATTTTTTAGTTCATTATATAAATCTAATGATTT 525719

Qy       1017 TTTAGCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGG 1076

Db     525718 TTTTTGTTTTATTTGTTTTTTAATACTATCAAATACTTCATATATTTTTTTTTTGTGTCCC 525659

Qy       1077 AATTGGTAAATCTAAGCTTGTTGCAAATTTAAAT 1110

Db     525658 AATTTGAGTCTCATTTAATTTAACAATTAAAAAT 525625
```

RESULT 13
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 4.4%; Score 55.6; DB 2; Length 7218;
Best Local Similarity 5.3%; Pred. No. 0.0094;
Matches 22; Conservative 224; Mismatches 168; Indels 0; Gaps 0;

```
Qy        676 TGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATA 735
```

```
Db       1040 TGGCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1099

Qy        736 CCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAA 795

Db       1100 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1159

Qy        796 TTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATG 855

Db       1160 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1219

Qy        856 AAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATT 915

Db       1220 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1279

Qy        916 TCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTGGTCGTGATATATTATCAATATGGATA 975

Db       1280 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1339

Qy        976 AACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATT 1035

Db       1340 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1399

Qy       1036 TTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATTGGTAAATCT 1089

Db       1400 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCT 1453


RESULT 14
US-08-998-416-186
; Sequence 186, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  186:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  615 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1074RP
US-08-998-416-186

  Query Match             4.4%;  Score 55.4;  DB 3;  Length 615;
  Best Local Similarity   47.8%;  Pred. No. 0.0064;
  Matches  223;  Conservative    0;  Mismatches  241;  Indels    3;  Gaps    2;

Qy        230 AGAGAATTCTTAGTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTT 289

Db         78 AAACTATTATTATCATTATTTAATAAATTAATTATTTGATTATTAATACTTATTATATAA 137

Qy        290 CAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCT 349

Db        138 TTATTATATAATTTACTTAATTCATCATTATTAATATTTATATAATTATAAAAATAATAT 197

Qy        350 TCGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTC 409

Db        198 TTAATATGAATACTATTTAGTCTATGTTCAAATTTTAAATTAGTTATTAAAATATTATT- 256

Qy        410 AGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCAT 469

Db        257 -AGATATTATTATTTTCTTTAATAAATTATTAAATAGATTATCAATAATTAATATATTAT 315

Qy        470 GTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACG 529

Db        316 TTATTAATTGTTT-ATTAAAATAATATATTTTATTATTATAAAGATTTAATTTATTTAAA 374

Qy        530 GAAATGCACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATA 589

Db        375 TATTGTAAATTATTATTTTTATTATAATATCTATTTTTATAAATATTATGTTGATTTATA 434

Qy        590 TTCTTGTATGTCTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGA 649

Db        435 TTATTTAACTTTTTATAAGAATTATTATTAAAATTAATTTAACTTTAATTTCTTATTAT 494

Qy        650 TTGTTAATTTGTTAAATGAGTCAAAATGGATGTTTCTTCAATTAACT 696

Db        495 TAATTTTTATATTATTTAATAAATTATATTCATTTTATTTATTTATT 541


RESULT 15
US-09-949-016-14624
; Sequence 14624, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14624
;   LENGTH: 147382
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(147382)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14624

  Query Match             4.3%;  Score 54.4;  DB 3;  Length 147382;
```

Best Local Similarity 44.6%; Pred. No. 0.031;
Matches 214; Conservative 0; Mismatches 266; Indels 0; Gaps 0;

```
Qy        280 TATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGT 339
              | || || |  |  | |  ||||     ||    |||  || | | ||  |||||| | |
Db     136697 TTTATTTATATATATTTATTTATATATATTTATTTATATATATTTATTTATATATTTATT 136756

Qy        340 ATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGT 399
              |  ||  ||  || |    |   |||  | ||  |   ||||        |||  ||  |
Db     136757 TATATTTATTTTTATATATTTATTTATATATATTTGTATATATTTATTTATATATTTATT 136816

Qy        400 GGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTT 459
              | | |||    |  |   ||  |   |||||        ||       | |  | |  |
Db     136817 TATATATTTATTTATATATTTATATATATTAGTTCATATATATTTATATATTTATATATA 136876

Qy        460 AATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATAT 519
              ||   || | ||| |  | || ||| |     |          || |    || |   |
Db     136877 TTAGTTATATATATTTATATATTAGTTATATATATTTATATATATTTTTATATATTTATT 136936

Qy        520 GGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATG 579
              || || | | ||     | ||| | |       |     ||||    ||||  |||
Db     136937 TATATATATGTATATATTAACATATTTATATATATTCATATATTTATATATAAATATATA 136996

Qy        580 CTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAAT 639
              ||||||||  | ||| | |  ||||||      |||||     | |  ||  | |
Db     136997 TATTAATATATTTATATATATTTATATATTAATATATGTTTATATATTTATACATATTTA 137056

Qy        640 AGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGT 699
              |  |   ||  ||  ||  || ||   |  | ||| || | | |  | |||   |  |
Db     137057 TGTATATTAATATATATATTTATATATATTTATATATGTATTTATATATATTTATATATT 137116

Qy        700 GGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAA 759
              | |   |  | ||| |       |  | |||   ||   ||||| | |   || |
Db     137117 TATATTTATTTATATATGTATTTATACATATTTATATATTTATATTTATTTATTTATATA 137176
```

Search completed: May 11, 2006, 05:23:22
Job time : 180.589 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:01:11 ; Search time 835.597 Seconds
(without alignments)
12528.801 Million cell updates/sec

Title: US-09-423-093B-1_COPY_8646_9911
Perfect score: 1266
Sequence: 1 atggtattaacagtgaaaaa.....gagcgaaaaatgtctattga 1266

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1266	100.0	14516	3	US-09-423-093A-1	Sequence 1, Appli
	2	96	7.6	3683	8	US-10-473-126-339	Sequence 339, App
	3	92.4	7.3	6292	7	US-10-221-714A-461	Sequence 461, App
	4	91.8	7.3	3683	8	US-10-473-126-193	Sequence 193, App
	5	90.2	7.1	6668	6	US-10-311-455-1670	Sequence 1670, Ap
c	6	86.4	6.8	1300	7	US-10-668-749A-1	Sequence 1, Appli
	7	82.8	6.5	1243	8	US-10-425-115-172717	Sequence 172717,
	8	82.2	6.5	7442	7	US-10-221-714A-409	Sequence 409, App
	9	80.6	6.4	3673778	6	US-10-312-841-1	Sequence 1, Appli
	10	80.4	6.4	6314	6	US-10-240-452-14	Sequence 14, Appl
	11	78.8	6.2	921	8	US-10-425-115-38710	Sequence 38710, A
	12	78	6.2	14006	6	US-10-311-455-1931	Sequence 1931, Ap
	13	76.4	6.0	5963	6	US-10-240-485-166	Sequence 166, App
	14	76.2	6.0	15479	7	US-10-257-166-45	Sequence 45, Appl
	15	75.4	6.0	5750	6	US-10-311-455-1981	Sequence 1981, Ap
	16	75.4	6.0	5750	7	US-10-221-714A-433	Sequence 433, App
	17	75.2	5.9	1004	8	US-10-425-115-81106	Sequence 81106, A
	18	74.2	5.9	1297	8	US-10-425-115-67687	Sequence 67687, A
	19	74.2	5.9	6113	6	US-10-311-455-1796	Sequence 1796, Ap
	20	74.2	5.9	15387	6	US-10-311-455-157	Sequence 157, App
	21	74	5.8	5145	6	US-10-311-455-321	Sequence 321, App
	22	74	5.8	5145	6	US-10-240-485-17	Sequence 17, Appl
	23	73.8	5.8	1039	8	US-10-425-115-58802	Sequence 58802, A
	24	73.8	5.8	1081	8	US-10-425-115-16756	Sequence 16756, A
	25	73.4	5.8	11555	6	US-10-311-455-589	Sequence 589, App
	26	73.2	5.8	5893	6	US-10-311-455-832	Sequence 832, App
	27	73	5.8	6767	7	US-10-221-714A-330	Sequence 330, App
	28	73	5.8	17131	6	US-10-311-455-1026	Sequence 1026, Ap
	29	72.2	5.7	16766	6	US-10-311-455-2130	Sequence 2130, Ap
	30	71.8	5.7	6192	7	US-10-221-714A-408	Sequence 408, App
	31	71.4	5.6	9539	5	US-10-239-676-52	Sequence 52, Appl
	32	71.4	5.6	9539	6	US-10-240-453-54	Sequence 54, Appl
	33	71.2	5.6	6291	7	US-10-221-714A-173	Sequence 173, App
	34	71	5.6	11812	5	US-10-239-676-210	Sequence 210, App
	35	71	5.6	11812	6	US-10-311-455-2092	Sequence 2092, Ap
	36	71	5.6	11812	6	US-10-240-453-306	Sequence 306, App
	37	71	5.6	11812	7	US-10-221-714A-468	Sequence 468, App
	38	71	5.6	17594	6	US-10-311-455-1999	Sequence 1999, Ap
	39	70.6	5.6	1062	8	US-10-425-115-120013	Sequence 120013,
	40	70.4	5.6	6079	6	US-10-311-455-231	Sequence 231, App
	41	70.2	5.5	8079	7	US-10-240-589C-122	Sequence 122, App
	42	70.2	5.5	11155	6	US-10-311-455-577	Sequence 577, App
	43	70.2	5.5	3673778	6	US-10-312-841-2	Sequence 2, Appli
	44	69.8	5.5	975	8	US-10-363-345A-16099	Sequence 16099, A
c	45	69.8	5.5	975	8	US-10-363-345A-16100	Sequence 16100, A

ALIGNMENTS

RESULT 1
US-09-423-093A-1
; Sequence 1, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 14516
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-423-093A-1

Query Match 100.0%; Score 1266; DB 3; Length 14516;
Best Local Similarity 100.0%; Pred. No. 4.3e-232;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8646 ATGGTATTAACAGTGAAAAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCACCGGTT 8705

Qy         61 ATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8706 ATTGAACAGTTTGTCAATCCAATTTGCATCTTCATTATCACACCACTAATACTCAACCAC 8765

Qy        121 CTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8766 CTGGGTAAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAG 8825

Qy        181 TTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8826 TTAATATGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTT 8885

Qy        241 AGTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8886 AGTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATT 8945
```

```
Qy       301 ATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCG 360
Db      8946 ATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCG 9005

Qy       361 AGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGAT 420
Db      9006 AGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGAT 9065

Qy       421 AATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAA 480
Db      9066 AATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAA 9125

Qy       481 GTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTC 540
Db      9126 GTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTC 9185

Qy       541 TTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGT 600
Db      9186 TTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGT 9245

Qy       601 CTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTG 660
Db      9246 CTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTG 9305

Qy       661 TTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTT 720
Db      9306 TTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTT 9365

Qy       721 GATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGC 780
Db      9366 GATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGC 9425

Qy       781 CTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCA 840
Db      9426 CTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCA 9485

Qy       841 ATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTG 900
Db      9486 ATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTG 9545

Qy       901 CTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATA 960
Db      9546 CTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTTGGTCGTGATATA 9605

Qy       961 TTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTA 1020
Db      9606 TTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAATTTTA 9665

Qy      1021 GCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 1080
Db      9666 GCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGGAATT 9725

Qy      1081 GGTAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCA 1140
Db      9726 GGTAAATCTAAGCTTGTTGCAAATTTAAATCTGGTTGCAGGGCTCGCACTTGCTGCTTCA 9785

Qy      1141 ACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCG 1200
Db      9786 ACGTTAATCGCAGCTCATTATGGCCTTTATGCAATATCTATGGTAAAAATAATATATCCG 9845

Qy      1201 GCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTC 1260
Db      9846 GCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAATAGAGCGAAAAATGTC 9905

Qy      1261 TATTGA 1266
Db      9906 TATTGA 9911
```

RESULT 2
US-10-473-126-339
; Sequence 339, Application US/10473126

; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 339
; LENGTH: 3683
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-339

Query Match 7.6%; Score 96; DB 8; Length 3683;
Best Local Similarity 45.3%; Pred. No. 2.7e-08;
Matches 428; Conservative 0; Mismatches 510; Indels 6; Gaps 2;

```
Qy       141 TAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTGGAGGATGTTC 200
Db      2574 TAGTTAGTTTTTTGTTTTTTTATTTGTTTATTGTATTTTTTTTTTTTATTTTGTTTTTAT 2633

Qy       201 CGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGATTTATCAAAAAAAAA 260
Db      2634 TTTATTGTTTGGTGGGAAAGGTTGTAGTTATTATGTATTTATTTATTTTTTATTTTTTATT 2693

Qy       261 TGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGAT 320
Db      2694 TTTTATTTTTTGATTTTTTTTTTATATTTTTATTTTTTTTTT--TATTTTTGTTTTTTT 2751

Qy       321 TTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTT 380
Db      2752 TTTTTTTTTTTTTTTAATTTTTTGTTTTTTTTATATATTTTATTTTTTTTTTAATTTTT 2811

Qy       381 ATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCT 440
Db      2812 GTTTTTTTTTATTTTTATTTTTATTTTGTTTTTTTTTTTTTTATTTTTGTTTTGGTTTGTT 2871

Qy       441 AAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTG 500
Db      2872 TTTTGTTTTTTTTTTTTTATTTTTATTTTTTTTTTTTTTTTGTTTTTTTTTTTTTATTTT 2931

Qy       501 GGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGC 560
Db      2932 TATTTTTTTTTTGTTTTGTTTATTTTTTGTTATTTTTATTTTTTTTTTTATTTTGTTTTTT 2991

Qy       561 CTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTT 620
Db      2992 TTTTTTTATTTTATTTTTTTTTTATTTTTTGTTTTTGTTTTTAATTTTTATGTTTTTATT 3051

Qy       621 CATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGAT 680
Db      3052 ATTTTTTGTTTTTTTTTTTTATTTTGATTTTTTATTTTATTTTGTGTTTTTTTTTTTAT 3111

Qy       681 GTTTCTTCA----ATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATAC 736
Db      3112 TTTTTTTTATTGTTTTTTTTTTTTGGTTTTTTATTTTGTTTTTTTTTTTTGTATGTTTATTT 3171

Qy       737 CATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAAT 796
Db      3172 TTTTTTTATTTTGTTTTTTTTTTTTTATGTTTATTTTTTTTTTATTTTGTTTTTTTTTTT 3231

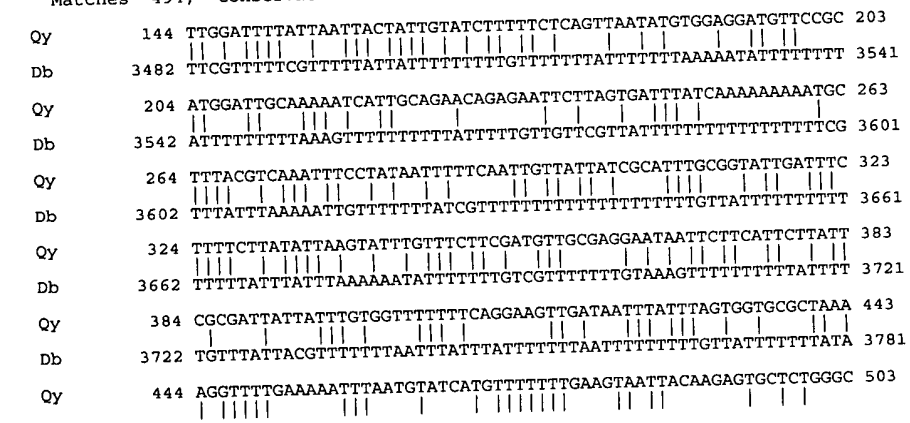
Qy       797 TGATGTTCACTCTTTCTGCGTCTGCAAATCAAAATATTACTACCAATGTTTGCTAGAATGA 856
Db      3232 TTATGTTTATTTTTTTTTTTATTTTGTTTTTTTTTTTTTTATGTTTATTTTTTTTTTTATTTT 3291

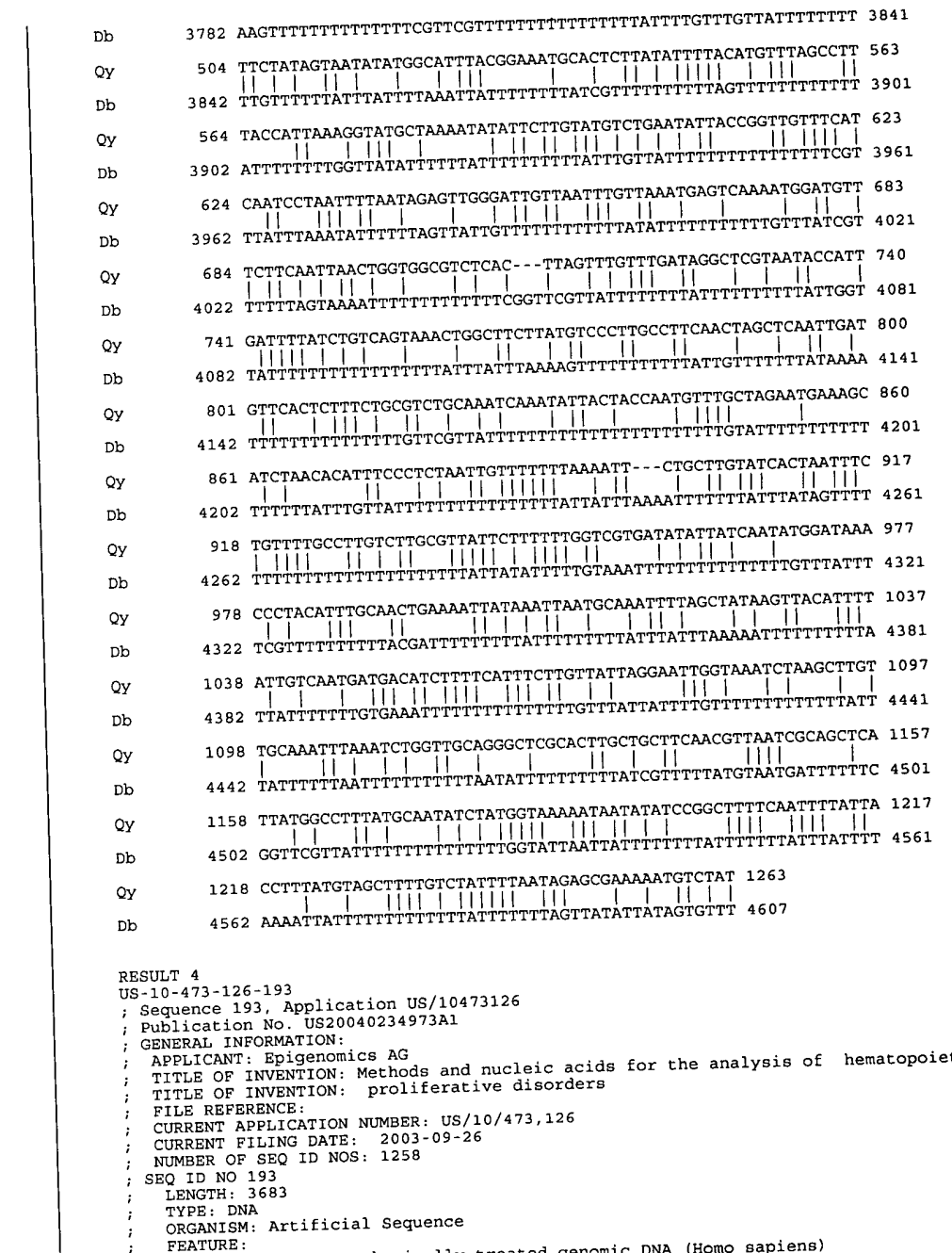
Qy       857 AAGCATCTAACACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTT 916
Db      3292 GTTTTTTTTTTTTTATTTTTAATTTTTTTTTTGTTTTTTTTTTTTTATTTTTATTTTTTT 3351
```



RESULT 3
US-10-221-714A-461
; Sequence 461, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 461
; LENGTH: 6292
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-461

Query Match 7.3%; Score 92.4; DB 7; Length 6292;
Best Local Similarity 43.9%; Pred. No. 1.6e-07;
Matches 494; Conservative 0; Mismatches 626; Indels 6; Gaps 2;

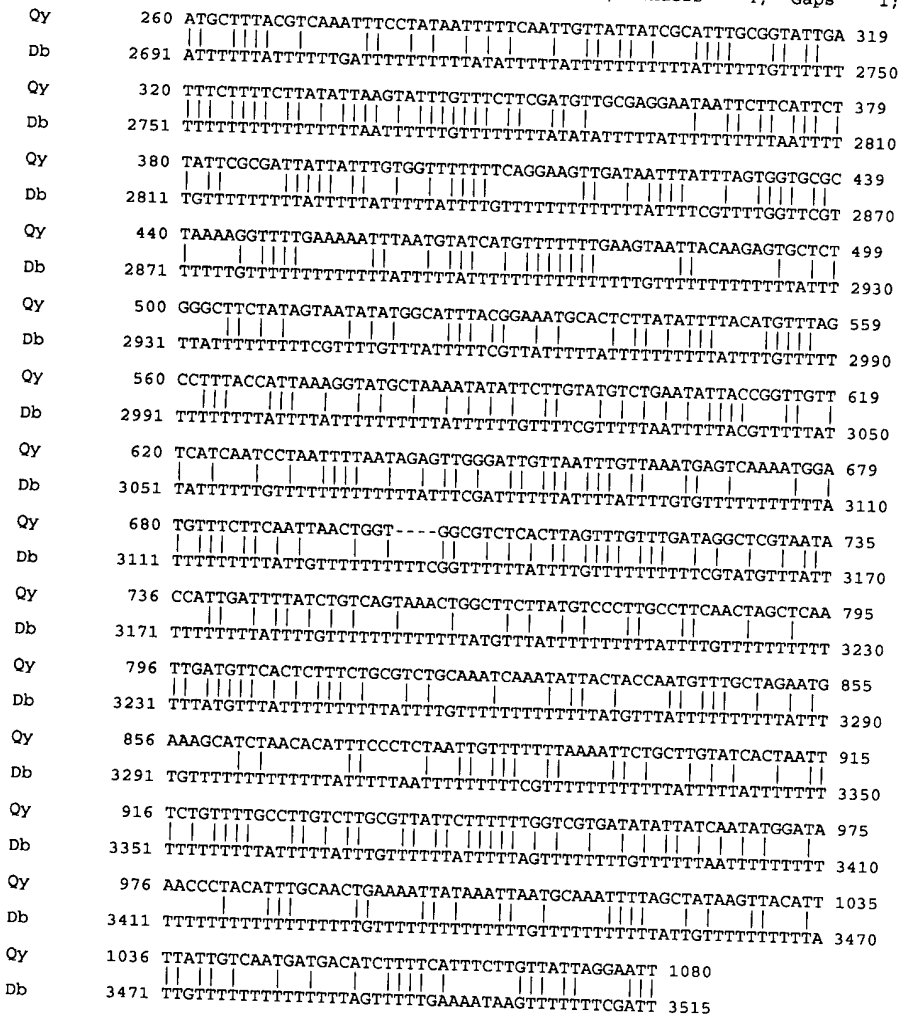




RESULT 4
US-10-473-126-193
; Sequence 193, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 193
; LENGTH: 3683
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

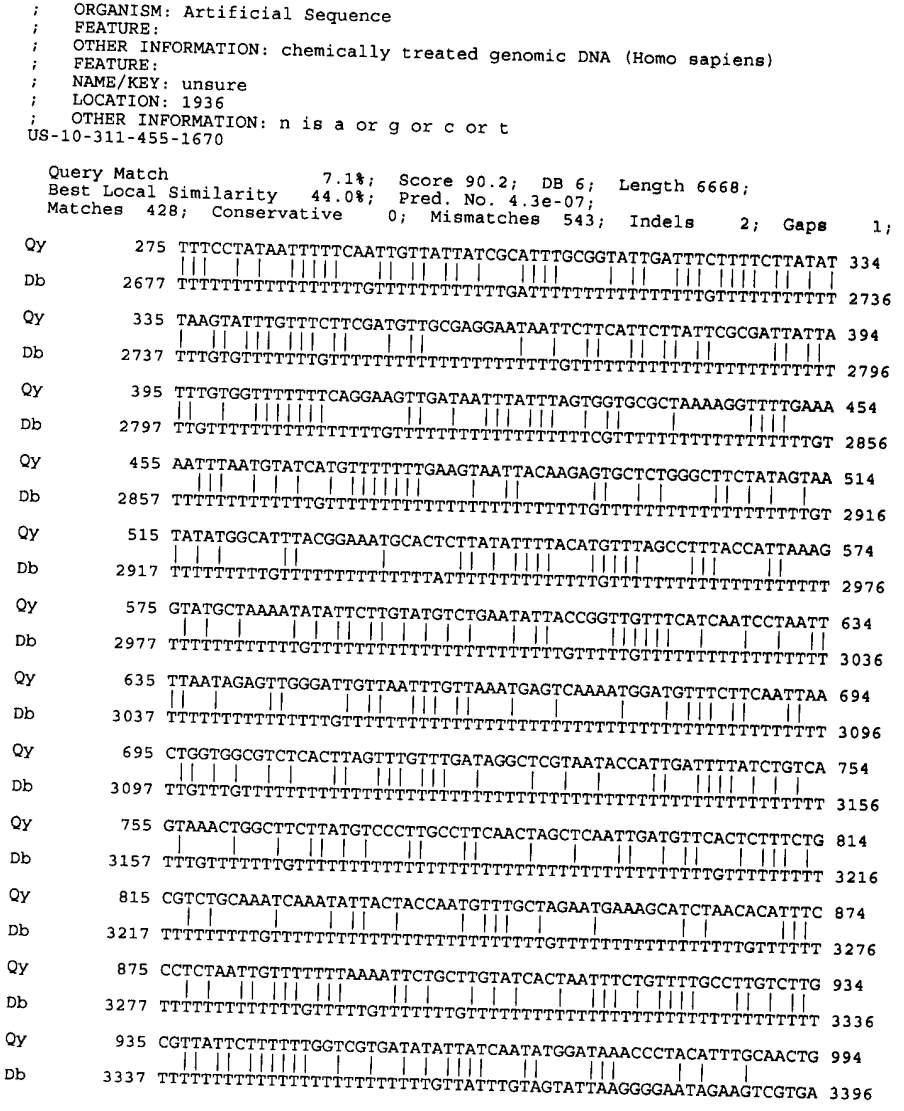
US-10-473-126-193

Query Match 7.3%; Score 91.8; DB 8; Length 3683;
Best Local Similarity 45.3%; Pred. No. 1.7e-07;
Matches 374; Conservative 0; Mismatches 447; Indels 4; Gaps 1;



RESULT 5
US-10-311-455-1670
; Sequence 1670, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1670
; LENGTH: 6668
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 1936
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1670

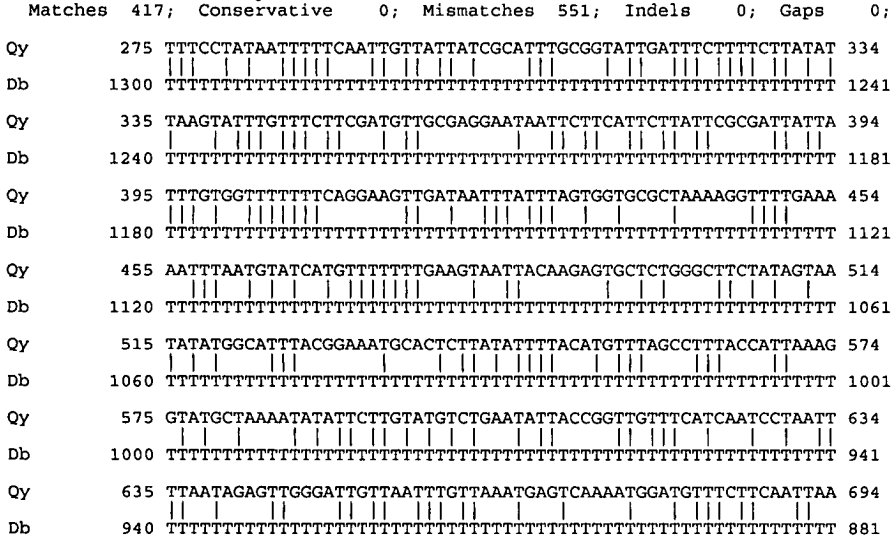
Query Match 7.1%; Score 90.2; DB 6; Length 6668;
Best Local Similarity 44.0%; Pred. No. 4.3e-07;
Matches 428; Conservative 0; Mismatches 543; Indels 2; Gaps 1;





RESULT 6
US-10-668-749A-1/c
; Sequence 1, Application US/10668749A
; Publication No. US20040110205A1
; GENERAL INFORMATION:
; APPLICANT: Agilent Technologies
; TITLE OF INVENTION: Methods and Systems for Nanopore Data Analysis
; FILE REFERENCE: 50112-1580
; CURRENT APPLICATION NUMBER: US/10/668,749A
; CURRENT FILING DATE: 2003-09-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
US-10-668-749A-1

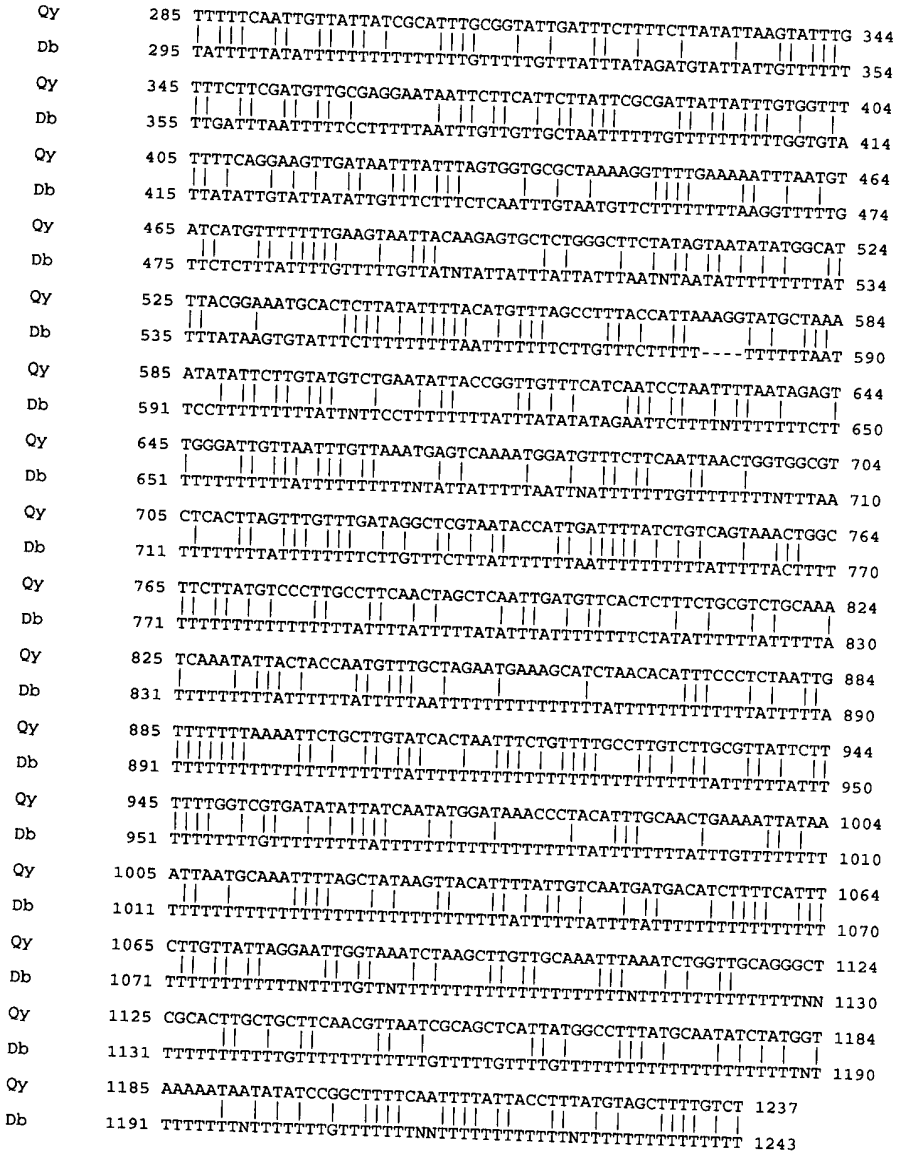
Query Match 6.8%; Score 86.4; DB 7; Length 1300;
Best Local Similarity 43.1%; Pred. No. 1.3e-06;
Matches 417; Conservative 0; Mismatches 551; Indels 0; Gaps 0;





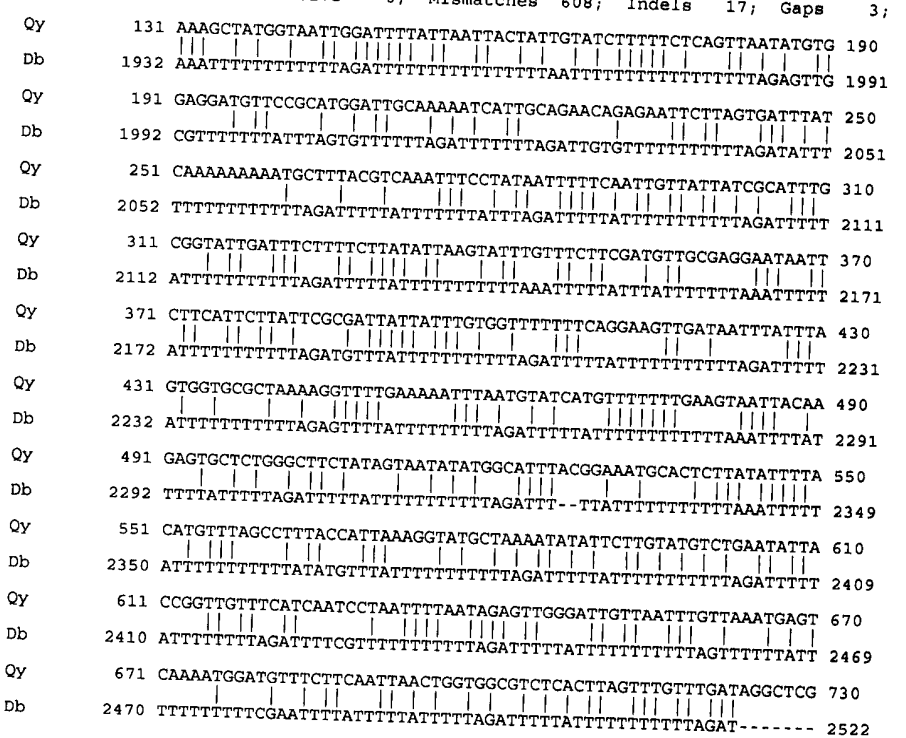
RESULT 7
US-10-425-115-172717
; Sequence 172717, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 172717
; LENGTH: 1243
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1243)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_89100C.1
US-10-425-115-172717

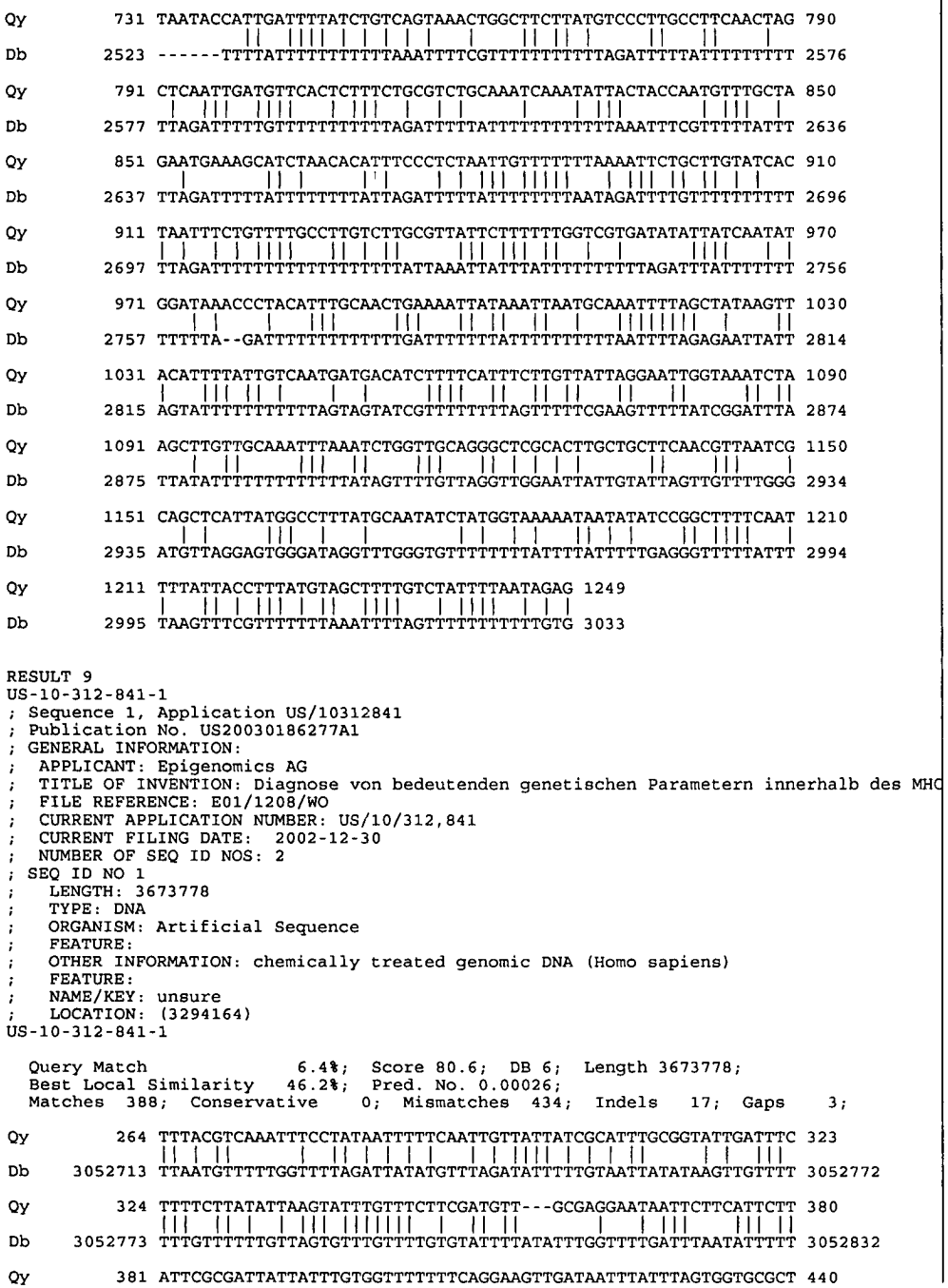
Query Match 6.5%; Score 82.8; DB 8; Length 1243;
Best Local Similarity 43.0%; Pred. No. 6.4e-06;
Matches 410; Conservative 0; Mismatches 539; Indels 4; Gaps 1;



RESULT 8
US-10-221-714A-409
; Sequence 409, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 409
; LENGTH: 7442
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 1792, 4861..4862
; OTHER INFORMATION: n is a or g or c or t
US-10-221-714A-409

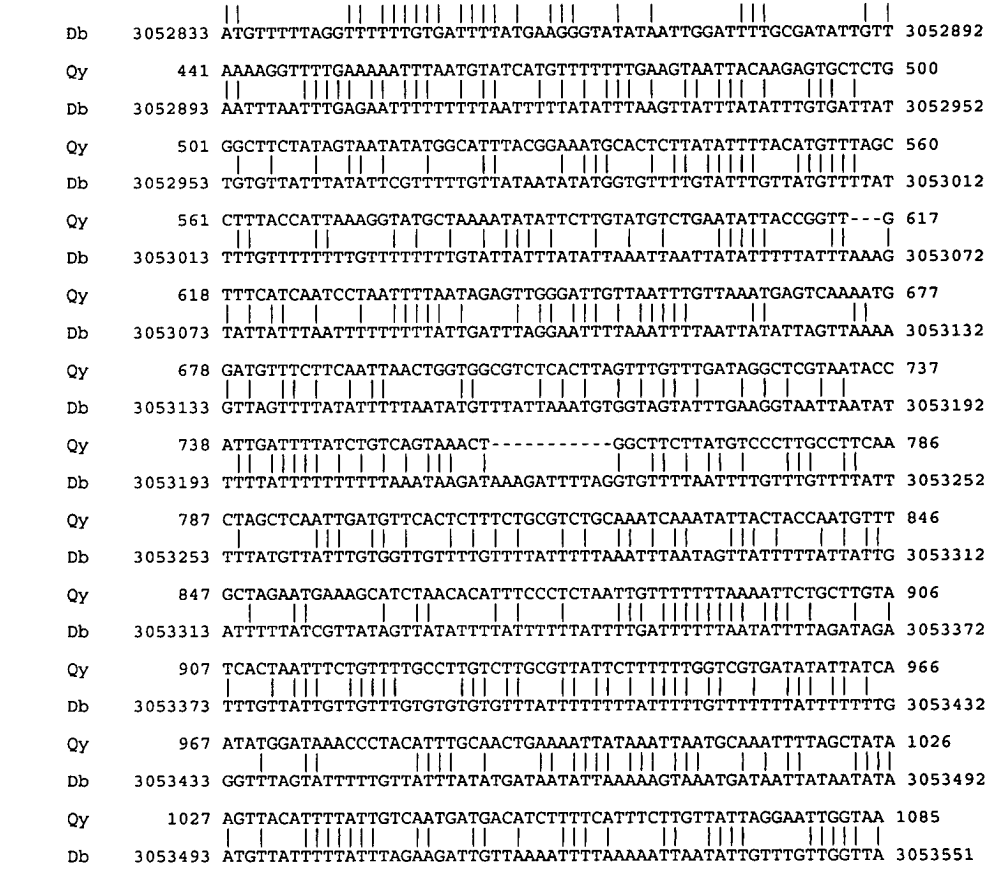
Query Match 6.5%; Score 82.2; DB 7; Length 7442;
Best Local Similarity 44.1%; Pred. No. 1.5e-05;
Matches 494; Conservative 0; Mismatches 608; Indels 17; Gaps 3;





RESULT 9
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

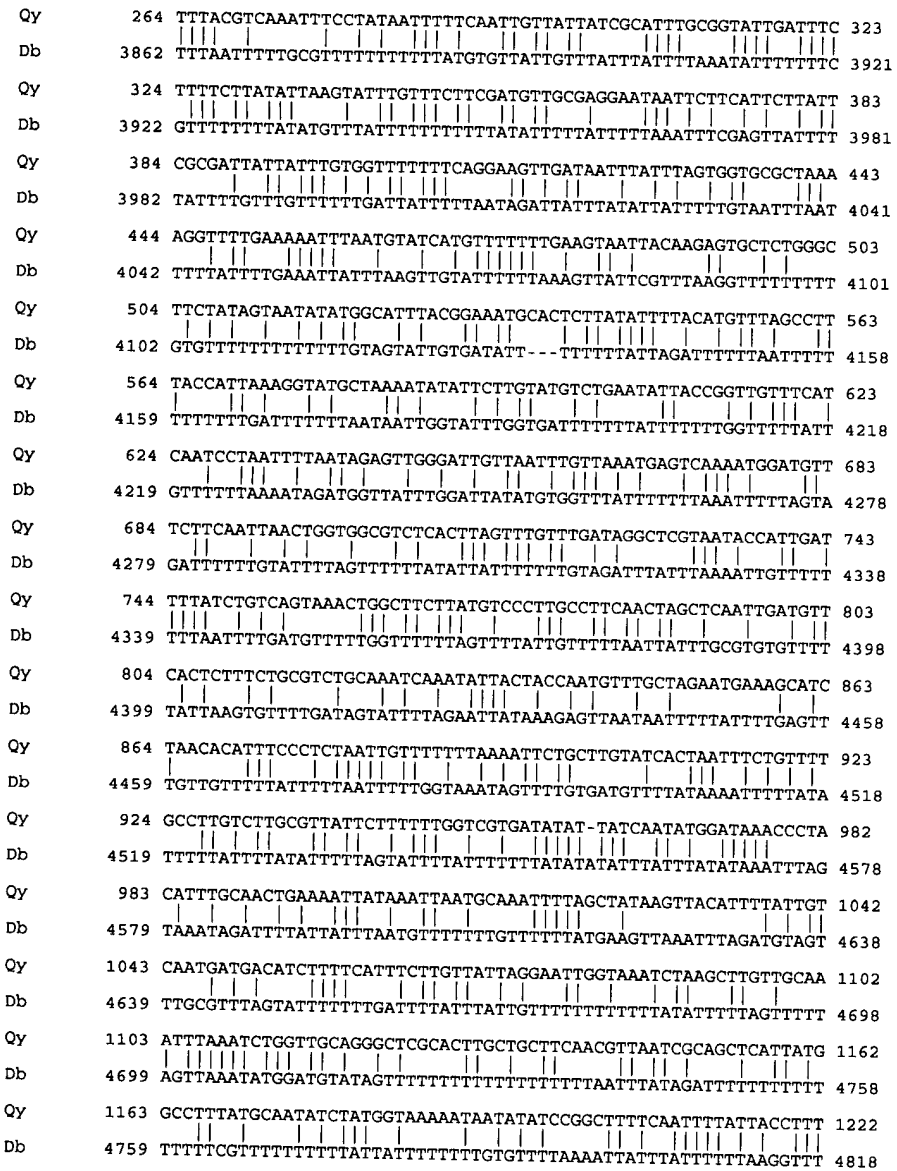
Query Match 6.4%; Score 80.6; DB 6; Length 3673778;
Best Local Similarity 46.2%; Pred. No. 0.00026;
Matches 388; Conservative 0; Mismatches 434; Indels 17; Gaps 3;



RESULT 10
US-10-240-452-14
; Sequence 14, Application US/10240452
; Publication No. US20030162194A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Apoptosis
; FILE REFERENCE: 5013.1006
; CURRENT APPLICATION NUMBER: US/10/240,452
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03969
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 78
; SEQ ID NO 14
; LENGTH: 6314

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-452-14

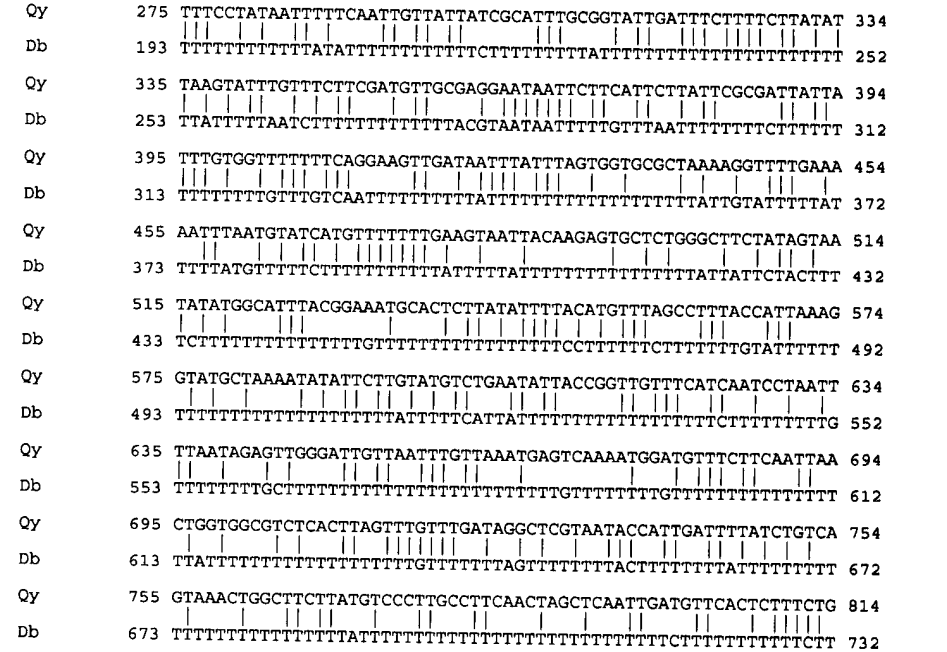
Query Match 6.4%; Score 80.4; DB 6; Length 6314;
Best Local Similarity 44.0%; Pred. No. 3.2e-05;
Matches 432; Conservative 0; Mismatches 546; Indels 4; Gaps 2;

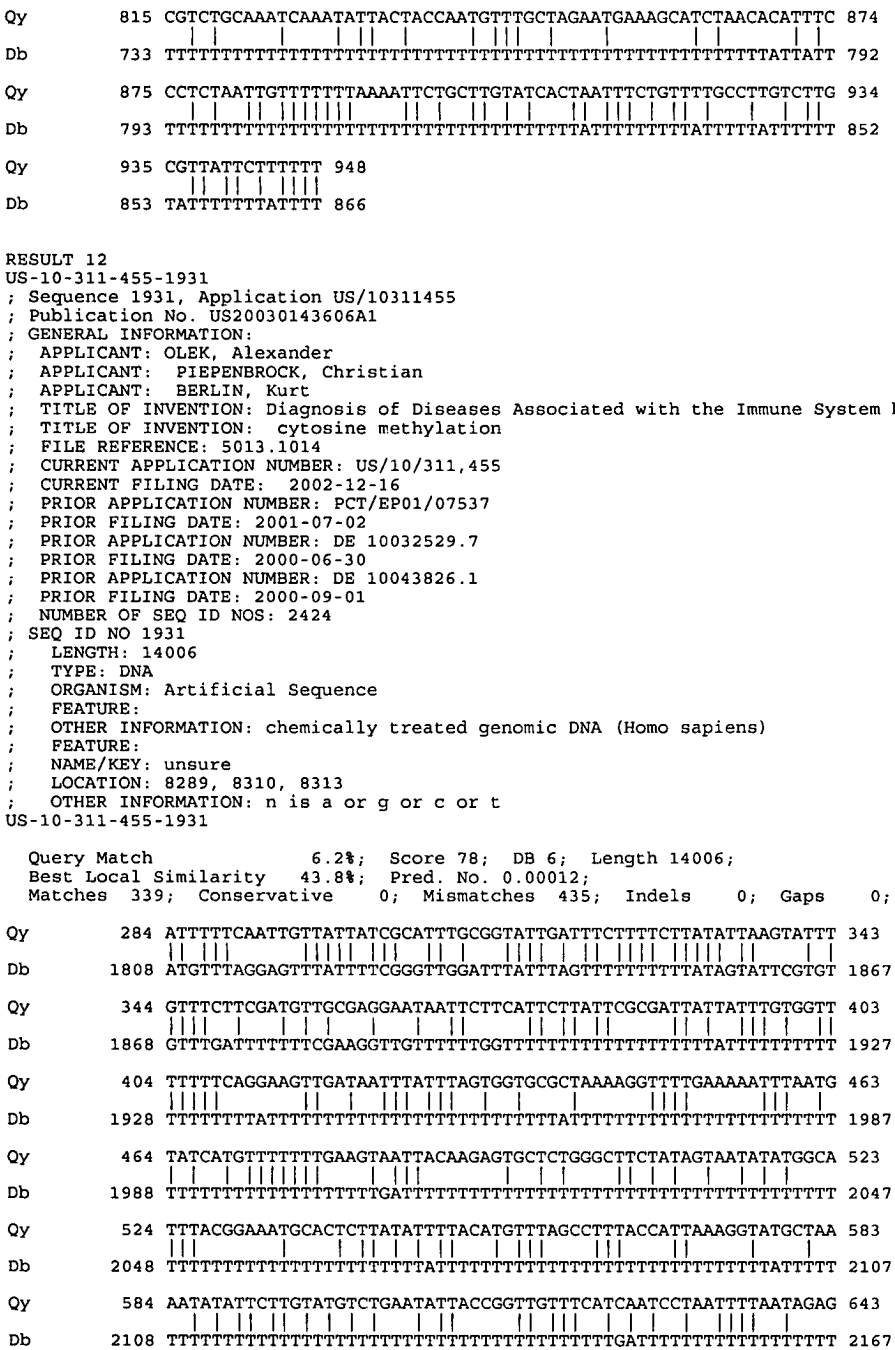


Qy 1223 ATGTAGCTTTTGTCTATTTTAA 1244
Db 4819 AATTTTGATATTTTTTTTAGAA 4840

RESULT 11
US-10-425-115-38710
; Sequence 38710, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 38710
; LENGTH: 921
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(921)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_135307C.1
US-10-425-115-38710

Query Match 6.2%; Score 78.8; DB 8; Length 921;
Best Local Similarity 44.8%; Pred. No. 3.3e-05;
Matches 302; Conservative 0; Mismatches 372; Indels 0; Gaps 0;





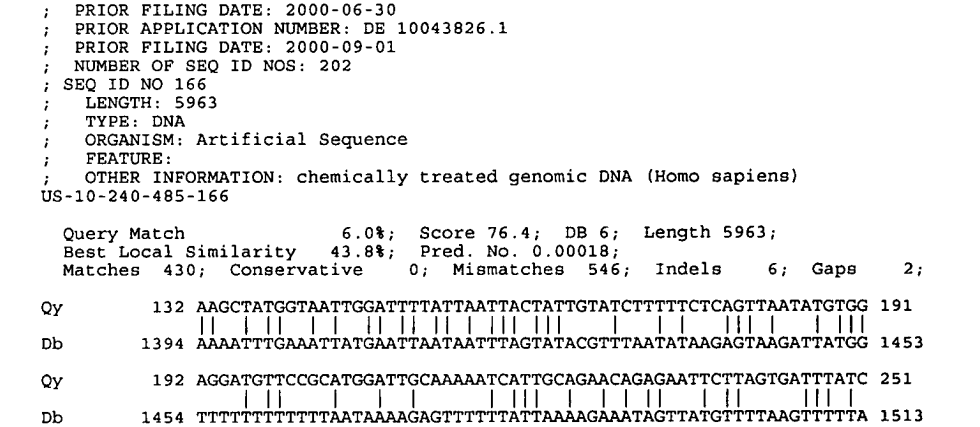
RESULT 12
US-10-311-455-1931
; Sequence 1931, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1931
; LENGTH: 14006
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 8289, 8310, 8313
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1931

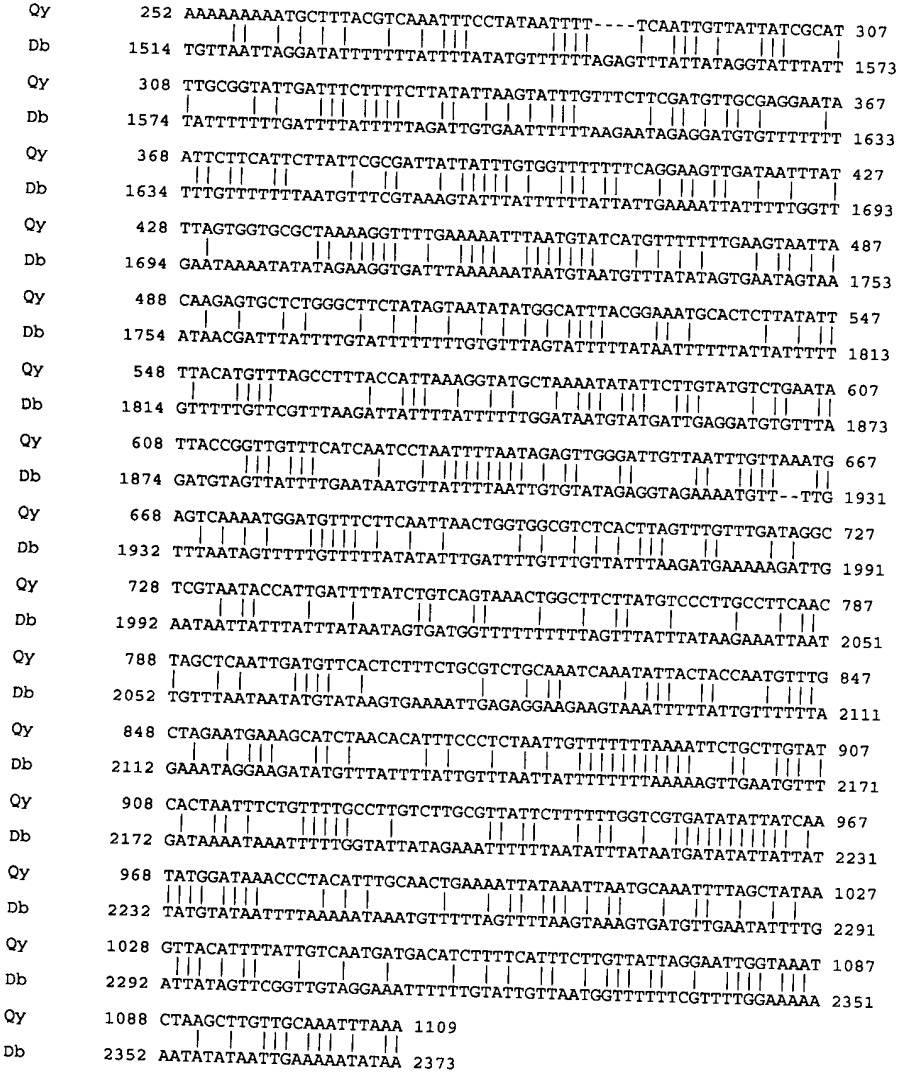
Query Match 6.2%; Score 78; DB 6; Length 14006;
Best Local Similarity 43.8%; Pred. No. 0.00012;
Matches 339; Conservative 0; Mismatches 435; Indels 0; Gaps 0;



RESULT 13
US-10-240-485-166
; Sequence 166, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: Metastasis
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240,485
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 166
; LENGTH: 5963
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-166

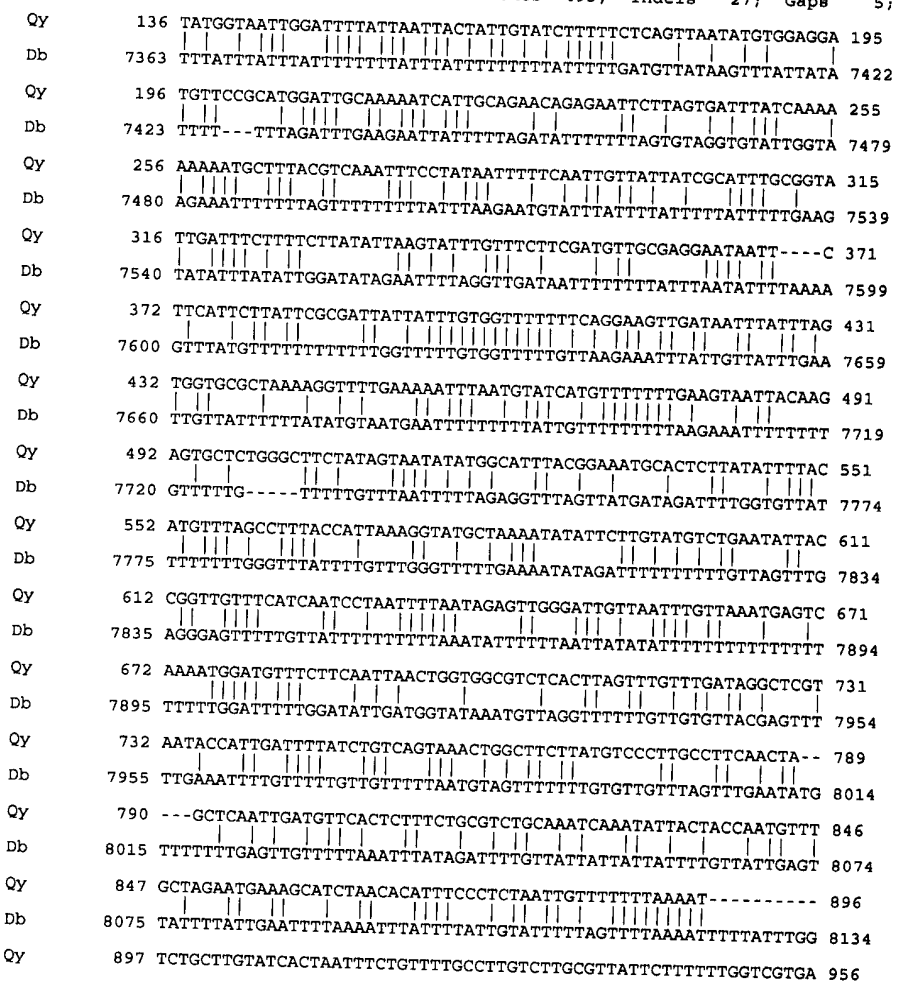
Query Match 6.0%; Score 76.4; DB 6; Length 5963;
Best Local Similarity 43.8%; Pred. No. 0.00018;
Matches 430; Conservative 0; Mismatches 546; Indels 6; Gaps 2;





RESULT 14
US-10-257-166-45
; Sequence 45, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 45
; LENGTH: 15479
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-45

Query Match 6.0%; Score 76.2; DB 7; Length 15479;
Best Local Similarity 46.3%; Pred. No. 0.00028;
Matches 449; Conservative 0; Mismatches 493; Indels 27; Gaps 5;



```
                | | || |||   | |||| | | ||    ||| ||   || || ||| ||    |
Db       8135 TTTTTTTTTATTTTTTATTTTTTTGTTAAAATGTTTTAATTTTTTGTTTGTTTTGAGATT 8194

Qy        957 TATATTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTAATGCAAAT 1016
              |  ||   ||  | | |          ||        | |||| | ||||| |  | ||
Db       8195 ATTTATAGTTATTGGAAAAGTTTTATAATAGTTGTTTTTAAATATTTAATTATTTTATAT 8254

Qy       1017 TTTAGCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTGTTATTAGG 1076
              |||   |||  ||   |||||||||||   || | |  | |||   |    | || ||
Db       8255 TTTTAATATTTGTGTTATTTTATTGTTGGTGTTAATTTATTTAATTGATGTTTTTTTGAT 8314

Qy       1077 AATTGGTAA 1085
                |||||||
Db       8315 TTTTGGTAA 8323


RESULT 15
US-10-311-455-1981
; Sequence 1981, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1981
;   LENGTH: 5750
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1981

  Query Match             6.0%;  Score 75.4;  DB 6;  Length 5750;
  Best Local Similarity   44.8%;  Pred. No. 0.00028;
  Matches  374;  Conservative    0;  Mismatches  456;  Indels    5;  Gaps    2;

Qy        280 TATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGT 339
              | ||||  |      |||  | |    ||||  ||| |  ||| |||| || | ||   |
Db       3547 TTTAATGATGGGTAGGTTTATTTATAGTTTGGAGTAGTCGTTTTTTTTTTTTTTTTTTT 3606

Qy        340 ATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGT 399
              |||| ||| ||   |       |    |  || ||  || |||||     || ||||||
Db       3607 TTTTTTTTTTTTTTTAGAAATATTTTTTTTTTTTTTTTTTATTTTTTTTTTTTATTTT 3666

Qy        400 GGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTT 459
              |||||||  |      ||  |  |||||||  | ||     |       ||||    | |||
Db       3667 TTTTTTTTTTATTTTTTTTTTTTTTTATTTTTCGTTTTATTTTTTTTTTTTTTTATTTT 3726

Qy        460 AATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATAT 519
              | | |  | |||||||     |  ||       ||  | |    || | |  |  | | |
Db       3727 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTT 3786

Qy        520 GGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATG 579
              |||    |  |    | || | ||||   | |||    |||    ||  |  | |
Db       3787 TTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATATTTT 3846

Qy        580 CTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAAT 639
              |    | | || ||||||  | |    | ||   | || |||  |   |  |  ||||  |
Db       3847 TTTTTTTTTTTTATTGTATTTTTTTTTTTTTTTGTTTTTTTGTTTTTTTTTTTTTTTT 3906

Qy        640 AGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGT 699

                ||    || ||  ||| ||   |   |     | |  || |  |    ||| |
Db       3907 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGGTTGGAGTGTAGTGGCGTAA-TTTC 3965

Qy        700 GGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAA 759
              ||  | |        ||| |||  |||| |       |   ||   |||   | | ||||
Db       3966 GGTTTATTGTAATTTTTATTTTTTAGGTTTAAGCGATTTTTTTGTTTTAGTTTTAGTAGT 4025

Qy        760 CTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCT 819
              || ||      |   |  | | |   |  |  |  | |  | | ||  |     |
Db       4026 TGGGATTAAAGGGGTTTATCGTTATTATTTTTTGTTAATTTTTGTAATTTTAGTAGAGAT 4085

Qy        820 GCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCT 879
                |    |||       | | |    |  |||    |  | ||    ||||   | |
Db       4086 AGGTTTTTGTTATGTTGGATAGTTTAGTGTTAAATTTTTGATTTTAGATGATTTATTTAT 4145

Qy        880 AATTGTTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTA 939
              |||||||  |   ||  | || ||    | | ||  |    |    | | ||   ||
Db       4146 TTCGGTTTTTTAAGTGTTGGGATTATAGGTATGAGTTATCGCGTTTGGTTTTTTTTTTTT 4205

Qy        940 TTCTTTTTTGGTCGTGATATATTATCAATATGGATAAACCCTACATTTGCAACTGAAAAT 999
              || ||||||  |   || | |||   | |||||  |       | |||     | |  |
Db       4206 TTTTTTTTTTTTAAAGAAAAATTGAAATTATGGTTTTTTAAAAATTTT----TTTTAGTT 4261

Qy       1000 TATAAATTAATGCAAATTTTAGCTATAAGTTACATTTTATTGTCAATGATGACATCTTTT 1059
              | ||  || ||  || |       |      ||   || |  ||| | || ||   | ||||
Db       4262 TTTAGTTTTATCGAATTACGTTTTGGTTATTGATTTGTCGTGTGATTGTTGTGTTTTTTT 4321

Qy       1060 CATTTCTTGTTATTAGGAATTGGTAAATCTAAGCTTGTTGCAAATTTAAATCTGG 1114
                ||| |  || ||   | ||  |   | |    || ||    |||||  |  ||
Db       4322 TGTTTTTATTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTGTTATTTATGTTCGG 4376


Search completed: May 11, 2006, 07:09:21
Job time : 844.597 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:23:58 ; Search time 654.207 Seconds
(without alignments)
7892.621 Million cell updates/sec

Title: US-09-423-093B-1_COPY_8646_9911
Perfect score: 1266
Sequence: 1 atggtattaacagtgaaaaa.....gagcgaaaaatgtctattga 1266

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

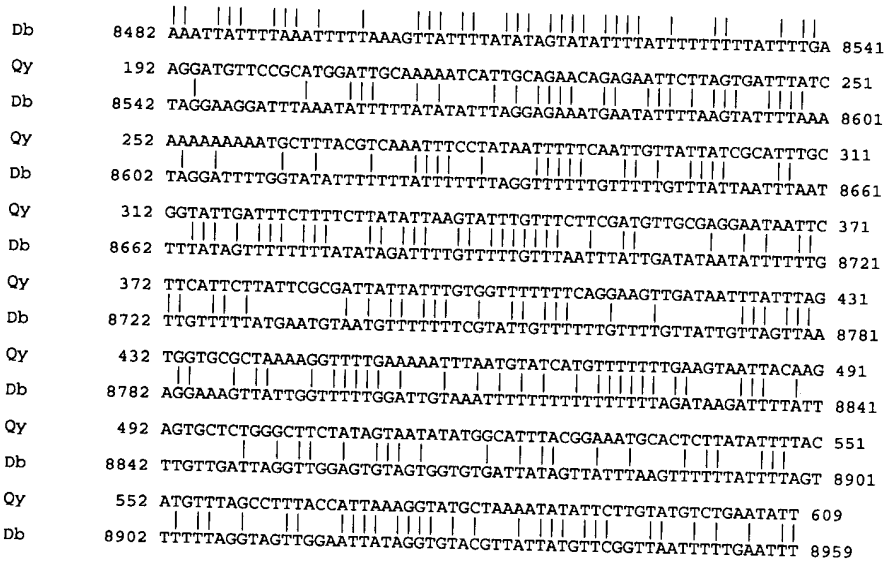
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	66.8	5.3	9347	10	US-10-240-708-36	Sequence 36, Appl
	2	59	4.7	6113	10	US-10-240-708-14	Sequence 14, Appl
	3	57	4.5	8391	11	US-10-932-182A-166179	Sequence 166179,
	4	57	4.5	8391	11	US-10-932-182A-166179	Sequence 166179,
	5	56.8	4.5	10467	10	US-10-240-708-2	Sequence 2, Appli
	6	56.8	4.5	19233	10	US-10-240-708-45	Sequence 45, Appl
	7	56	4.4	6306	10	US-10-240-708-50	Sequence 50, Appl
	8	55	4.3	5562	10	US-10-240-708-63	Sequence 63, Appl
c	9	55	4.3	201990	10	US-10-995-561-13303	Sequence 13303, A
	10	54.6	4.3	8961	10	US-10-240-708-80	Sequence 80, Appl
	11	54.2	4.3	9347	10	US-10-240-708-35	Sequence 35, Appl
	12	54.2	4.3	28536	17	US-11-011-332A-155	Sequence 155, App
	13	53.4	4.2	6499	17	US-11-011-332A-91	Sequence 91, Appl
c	14	52.8	4.2	4069	10	US-10-793-626-3976	Sequence 3976, Ap
	15	51.4	4.1	5152	10	US-10-240-708-73	Sequence 73, Appl
	16	51.4	4.1	6070	10	US-10-240-708-10	Sequence 10, Appl
	17	51.4	4.1	6456	17	US-11-011-332A-92	Sequence 92, Appl
	18	50.6	4.0	6156	10	US-10-240-708-60	Sequence 60, Appl
	19	50.6	4.0	6257	17	US-11-011-332A-95	Sequence 95, Appl
	20	50.4	4.0	5455	10	US-10-240-708-33	Sequence 33, Appl
	21	50.4	4.0	28536	17	US-11-011-332A-156	Sequence 156, App
	22	50.2	4.0	5666	10	US-10-240-708-29	Sequence 29, Appl
	23	50	3.9	6583	10	US-10-240-708-26	Sequence 26, Appl
c	24	49.8	3.9	305312	10	US-10-995-561-13236	Sequence 13236, A
	25	49.6	3.9	11049	10	US-10-240-708-21	Sequence 21, Appl
	26	49.4	3.9	396	18	US-11-250-759-53	Sequence 53, Appl
c	27	49.4	3.9	621	7	US-09-925-065A-874667	Sequence 874667,
	28	49.4	3.9	6499	17	US-11-011-332A-77	Sequence 77, Appl
	29	49.4	3.9	6669	10	US-10-240-708-6	Sequence 6, Appli
c	30	49.2	3.9	139054	17	US-11-121-086-96	Sequence 96, Appl
	31	49	3.9	6866	10	US-10-240-708-20	Sequence 20, Appl
	32	49	3.9	28536	17	US-11-011-332A-151	Sequence 151, App
	33	48.8	3.9	5535	10	US-10-240-708-18	Sequence 18, Appl
	34	48.6	3.8	8093	10	US-10-240-708-31	Sequence 31, Appl
	35	48.6	3.8	8607	10	US-10-240-708-72	Sequence 72, Appl
c	36	48.4	3.8	632	12	US-10-301-480-362422	Sequence 362422,
c	37	48.4	3.8	632	12	US-10-301-480-975831	Sequence 975831,
	38	48.4	3.8	994	12	US-10-301-480-565200	Sequence 565200,
	39	48.4	3.8	994	12	US-10-301-480-1178609	Sequence 1178609,
c	40	48.4	3.8	1184	7	US-09-925-065A-285428	Sequence 285428,
c	41	48.4	3.8	2067	11	US-10-932-182A-4341	Sequence 4341, Ap
c	42	48.4	3.8	2067	11	US-10-932-182A-4341	Sequence 4341, Ap
	43	48.4	3.8	6040	10	US-10-240-708-69	Sequence 69, Appl
	44	48.2	3.8	623	7	US-09-925-065A-580866	Sequence 580866,
	45	48.2	3.8	623	7	US-09-925-065A-580867	Sequence 580867,

ALIGNMENTS

RESULT 1
US-10-240-708-36
; Sequence 36, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 36
; LENGTH: 9347
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-36

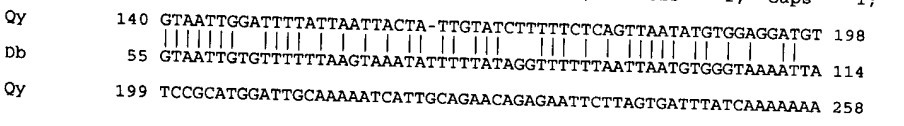
Query Match 5.3%; Score 66.8; DB 10; Length 9347;
Best Local Similarity 46.2%; Pred. No. 0.066;
Matches 221; Conservative 0; Mismatches 257; Indels 0; Gaps 0;

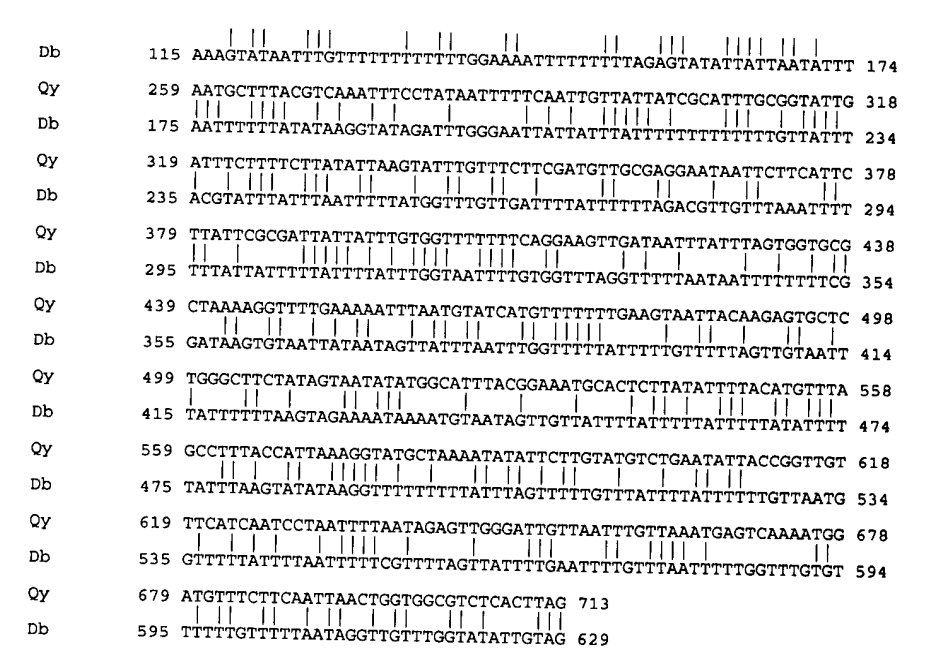
Qy 132 AAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTGG 191



RESULT 2
US-10-240-708-14
; Sequence 14, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 14
; LENGTH: 6113
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-14

Query Match 4.7%; Score 59; DB 10; Length 6113;
Best Local Similarity 45.0%; Pred. No. 1.2;
Matches 259; Conservative 0; Mismatches 315; Indels 1; Gaps 1;





RESULT 3
US-10-932-182A-166179
; Sequence 166179, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166179
; LENGTH: 8391
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166179

Query Match 4.5%; Score 57; DB 11; Length 8391;
Best Local Similarity 50.7%; Pred. No. 2.7;
Matches 191; Conservative 0; Mismatches 180; Indels 6; Gaps 2;



```
Qy        440 TAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTT----TTGAAGTAATTACAAGAGTG 495
              || ||   ||  | | |||| |  ||| | ||  |    |  |  || ||| | |||
Db       5472 TATAAATATTATATATATTTTAGATATAGTTTTAATTGAATATATATATTTAAAGGAGGA 5531

Qy        496 CTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTAT--ATTTTACAT 553
              | |    ||       ||||||  |   |||      | ||  | |  |||  ||| |  ||
Db       5532 ATATATTATTAAGATTTAATATTTTATATTATTATATATATAATGATATAGATTATTAAT 5591

Qy        554 GTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCG 613
               ||     ||  ||  |     |   |  || |||| || | |||    ||||||  |
Db       5592 TAATAAATCATATTATACATTTATTATCATTATTTATTATTATTTGTTGTTATATTAGGG 5651

Qy        614 GTTGTTTCATCAATCCT 630
              ||    |  ||| |||
Db       5652 ATTTGAACCCCAAACCT 5668


RESULT 4
US-10-932-182A-166179
; Sequence 166179, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAO, YOSHIHIRO
;  APPLICANT:  NAKAMURA, NORIHISA
;  APPLICANT:  KODAMA, YUKIKO
;  APPLICANT:  FUJIMURA, TOMOKO
;  APPLICANT:  ASHIKARI, TOSHIHIKO
;  TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
;  FILE REFERENCE: 030685-043
;  CURRENT APPLICATION NUMBER: US/10/932,182A
;  CURRENT FILING DATE:  2004-09-02
;  NUMBER OF SEQ ID NOS: 197023
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166179
;   LENGTH: 8391
;   TYPE: DNA
;   ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166179

  Query Match             4.5%;  Score 57;  DB 11;  Length 8391;
  Best Local Similarity   50.7%;  Pred. No. 2.7;
  Matches  191;  Conservative    0;  Mismatches  180;  Indels    6;  Gaps    2;

Qy        260 ATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGA 319
              ||   |||  |  ||| || |||||| | | ||||  ||| |||  |||        || |
Db       5292 ATATATTATTTATAATCTCTTATAATATATTAATTAATATGATCTTATTATATTCTTTTA 5351

Qy        320 TTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCT 379
              ||   | | |||||||    |||  ||  ||  ||  |   |     ||| || ||| |
Db       5352 TTAAATATATTATATTGTTAATTATTTATTTATATTATATTATATTATATTGTTAATTAT 5411

Qy        380 TATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGC 439
              |  |    |||||  | | |  || |||  |     || | ||| |||  ||  ||
Db       5412 TTATTTATATTATATTATTTAATTCTTTATATTTTATTAAGAATATATAAAGATGTTTTA 5471

Qy        440 TAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTT----TTGAAGTAATTACAAGAGTG 495
              || ||   ||  | | |||| |  ||| | ||  |    |  |  || ||| | |||
Db       5472 TATAAATATTATATATATTTTAGATATAGTTTTAATTGAATATATATATTTAAAGGAGGA 5531

Qy        496 CTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTAT--ATTTTACAT 553
              | |    ||       ||||||  |   |||      | ||  | |  |||  ||| |  ||
Db       5532 ATATATTATTAAGATTTAATATTTTATATTATTATATATATAATGATATAGATTATTAAT 5591

Qy        554 GTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCG 613
               ||     ||  ||  |     |   |  || |||| || | |||    ||||||  |
Db       5592 TAATAAATCATATTATACATTTATTATCATTATTTATTATTATTTGTTGTTATATTAGGG 5651

Qy        614 GTTGTTTCATCAATCCT 630
              ||    |  ||| |||
Db       5652 ATTTGAACCCCAAACCT 5668


RESULT 5
US-10-240-708-2
; Sequence 2, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/240,708
;  CURRENT FILING DATE:  2002-10-03
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
;   LENGTH: 10467
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-2

  Query Match             4.5%;  Score 56.8;  DB 10;  Length 10467;
  Best Local Similarity   43.3%;  Pred. No. 3;
  Matches  425;  Conservative    0;  Mismatches  547;  Indels   10;  Gaps    3;

Qy        149 TTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTGGAGGATGTTCCGCATGGA 208
              | | |  | ||  | || |   ||||  |  | |  |   ||   | || |       |
Db       1879 TATAAAAATTTTTTTTTTTGAATTTTATTTGGGTTTTCGTTGTGTGGTGGTTAATTGTGT 1938

Qy        209 TTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGATTTATCAAAAAAAAATGCTTTAC 268
              ||   |||    ||  ||   |  |     | |||  ||| |    |||||||  |||
Db       1939 TTAATAAATATGTTTTAGTTTATTGTGGATTGAGTATTTTTTTTTTAAAAAATTTTTTTT 1998

Qy        269 GTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTC 328
              |    |||  |   ||  |  | | || |  |    |||     | ||| ||| ||||
Db       1999 TTAGGGTTTGGTGATATAGTAAAGGTTTTTTATTTTTTTTTTTTTTTTGTTTTTTTTTT 2058

Qy        329 TTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGA 388
              |||| ||   || || |||  | | | |||      |   |||  |  ||  |||
Db       2059 TTATTTTTTATAATTTTTTAATAGTTATTGTATTAACGGATTTATTTTTTATATAATTAG 2118

Qy        389 TTATTAT--TTGTGGTTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGG 446
              || |||   || |     ||||        || ||||||||||  | |   | |  ||
Db       2119 TTTTTAAGGTTTTAAGGATTTTTTTTTTTTTTATAATTTATTTTTTAGGTTGTTTTAATT 2178

Qy        447 TTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTC 506
              ||     |||||||  ||   | ||     | |    |||  ||   |  | ||
Db       2179 ATTGTTTTAATTTAAATTAGTTTTTTGAAGTAAGAGTATTTAAATT-TATTATGTTTAGT 2237

Qy        507 TATAGTAATATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTTTAC 566
              ||| |  | | ||| | ||   ||  || | | |||  | |  |       |    ||
Db       2238 TATTATTTTGTGTGGGAGTTTAAGATTTGAATTTTTAATTATGTCGATAGTTTGTTTATT 2297

Qy        567 CATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATATTACCGGTTGTTTCATCAA 626
              |   | | |  |||  || ||  |     | |||||  |   |  ||||  | ||
Db       2298 TAAATATTTTGTTATAATTTAAATAGTATATATTTTGAATTAAATTAATAGTTTTTTTAA 2357

Qy        627 TCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATGAGTCAAAATGGATGTTTCT 686
              || ||||  |     | |  ||  |  ||  |   || | |     |  |  ||  |
Db       2358 ATTTAGTTTTTTTTTTTGTAGTTTTATTGTTAATGGTATTATTTTTTTTTTAGTTTATT 2417

Qy        687 TCAATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGCTCGTAATACCATTGATTTT 746
```

```
Db       2418 AAGGTTAGAAATTTGCGTTATTTTTGTTGTTCGTTTTTGGTTTTATTTTTTGGATTTTA 2477

Qy        747 ATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGCCTTCAACTAGCTCAATTGATGTTCAC 806

Db       2478 TTAGGTAGTTGTTAAGTTTTATTTTGTTTGTTATTTTTGTATTGTGTTTTAAATGTGTTT 2537

Qy        807 TCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTGCTAGAATGAAAGCATCTAA 866

Db       2538 TTTTTTAGTTTTAGTT-------TTTTATTATTTTTTATTGGGTTATTGTAAAGGTTTAT 2590

Qy        867 CACATTTCCCTCTAATTGTTTTTTTAAAATTCTGCTTGTATCACTAATTTCTGTTTTGCC 926

Db       2591 AATTTTTTTATTTTTTTATATGTTAATTTAGGTGGTATATTTTTTTTTATAGTTATTTAT 2650

Qy        927 TTGTCTTGCGTTATTCTTTTTTGGTCGTGATATATTATCAATATGGATAAACCCTACATT 986

Db       2651 TATAATTTTGATTTTTTTTTTATAGAGTTTTTTAAATGTTATTTGTTTTAGGAGATTTTT 2710

Qy        987 TGCAACTGAAAATTATAAATTAATGCAAATTTTAGCTATAAGTTACATTTTATTGTCAAT 1046

Db       2711 GATTTTTATTTTTTTTAGTTTTTGGGGTTTTTTTTTTTTGAGTTTTTTTTATTTAGTATT 2770

Qy       1047 GATGACATCTTTTCATTTCTTGTTATTAGGAATTGGTAAATCTAAGCTTGTTGCAAATTT 1106

Db       2771 TGTTAAATTATTTATTTTTTTTCGTATTAGAGTTATTGGATTTTTTTTTTTTTTATGGTA 2830

Qy       1107 AAATCTGGTTGCAGGGCTCGCA 1128

Db       2831 TATATATATTGTAGTGTTTGAA 2852


RESULT 6
US-10-240-708-45
; Sequence 45, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/240,708
;  CURRENT FILING DATE:  2002-10-03
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 45
;   LENGTH: 19233
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (76, 178..179, 273, 586, 648, 651, 920, 1014, 1173, 1197, 1228)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1264, 1267, 1559, 1607, 1677, 1703, 1742, 1848, 1850, 2033)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (2168, 2254, 2257, 2280, 2291, 2723, 3018, 3080, 4205, 4209)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (4216, 4225, 4239, 4246, 4254, 4265, 4270, 4274, 4288, 4294..4295)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (4298, 4363, 4610, 4627, 4675, 4693, 4698, 4738, 4777, 4781)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (4787, 4825, 4829, 4838, 4855, 5462, 5494, 5496, 5500, 5674)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (5690, 5697, 5705, 5715, 5890, 5910, 5913, 6027, 6080, 6109)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (6244, 6288, 6776, 6873, 6891, 6980, 7028, 7656, 7876, 8030)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (8059, 8075, 8632, 8636, 9700, 9717, 9815, 9818, 9867, 9879)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (10005, 10075, 10150, 10159, 10165, 10356, 11123, 11166, 11360)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (11703, 11933, 12050, 12340, 12981, 13564, 13764..13765, 13787)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (13846, 14168, 16566, 17006, 17217, 17650, 17653, 17656, 17898)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (18130, 18143, 18151, 18188, 18498, 18549, 18936, 18946, 19165)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (19200)
US-10-240-708-45

  Query Match             4.5%;  Score 56.8;  DB 10;  Length 19233;
  Best Local Similarity   47.5%;  Pred. No. 3.2;
  Matches  224;  Conservative    0;  Mismatches  246;  Indels    2;  Gaps    2;

Qy        230 AGAGAATTCTTAGTGATTTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTT 289

Db       5200 AAAATATGTTTATATTTTTTTTACGTAAGAATAAATTAAATAAACGTTTGTTTTTTTTTG 5259

Qy        290 CAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCT 349

Db       5260 TATTTTTTATTTGCGGATATTTTTTGTTTTTTTTTTTTTTTTTTATTTATAGTTAAATTTT 5319

Qy        350 TCGA-TGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTT 408

Db       5320 TTGTTTGGTTTTAGTTATTTTTTTTTTTCGTTTTTTATTAATTTTTTAATTTATTTTAGT 5379

Qy        409 CAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCA 468

Db       5380 TATTTTTTTGATCGTTGTAGAAAATTTTGGTTGAAGGTATTAGTGGTTTTTAAGTTGTTA 5439

Qy        469 TGTTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTAC 528

Db       5440 AATTTAGTTTTTTTTTCGTTTTTNAGTTTATTTGAGTTTT-TAGTAGCGTTTGCGTNTNTA 5498

Qy        529 GGAAATGCACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATAT 588

Db       5499 ANGTTTTGTTTGTTTTAGGTTATATTTTTTTTTTTGGTTTTTAGTGGTTTAGTTTGGTAG 5558

Qy        589 ATTCTTGTATGTCTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGG 648

Db       5559 GTTTATTTTCGTTTTATTTGTAAGTTTTTTTTTGTAAATTTTTGTTGAATTAAGGTTTTT 5618

Qy        649 ATTGTTAATTTGTTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTG 700

Db       5619 TTTAGGTTTTTTTGATTTTCGAGGTAATTTTTTTTTTTTTTTTTAGTTTGGG 5670


RESULT 7
US-10-240-708-50
; Sequence 50, Application US/10240708
```

; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 50
; LENGTH: 6306
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-50

Query Match 4.4%; Score 56; DB 10; Length 6306;
Best Local Similarity 43.4%; Pred. No. 3.8;
Matches 355; Conservative 0; Mismatches 460; Indels 3; Gaps 2;

```
Qy        291 AATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTT 350
              ||||   ||| | |   ||     |  ||  |||| | |      |||||| |||| ||
Db       4850 AATTTGGATTTTTGTTATTAGTAAAAAAATGATTTTTTTAAATGGTGGTATTAGTTTATT 4909

Qy        351 CGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCA 410
              |   |  |  | ||||    |   || |  ||   |  |||| || ||  ||||||
Db       4910 TTTTTATTTGGAGTATAA-GATAGTTTATAGTTATTGTATATTTTTGGTAATTTTTGGAG 4968

Qy        411 GGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAA--TGTATCA 468
              ||  | | ||   | |   ||   |       || |     | |||||  |||||
Db       4969 TTTTTTTTTTTAATTTGATGGATATGAAGTGTTTTTTTGTCGTGATTTTAATTTGTATTT 5028

Qy        469 TGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTAC 528
              | |  || |  |  | ||  || | |  | | |  || ||| |     |   |  | |
Db       5029 TTTGATTATTGAAAAGTTTGAATATTTTTTTAGATTTTTATTGGTTATTTATGTTTATTT 5088

Qy        529 GGAAATGCACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATAT 588
              |      || | ||||   |  ||    ||||  | |       |  | | ||||
Db       5089 TTTTTTTTTTGATTTTTTTTTTTTTTATTGATTTTTAAGAGTTTTTTATTAGTGAGATAT 5148

Qy        589 ATTCTTGTATGTCTGAATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGG 648
              |  ||           | ||     ||| || ||   | |  |||| |    ||
Db       5149 TGTAATGGTATAAGTGTTGTTTGTTTTTGGTTTATTTTTTCGTTTTTTTTTTTTTATTTTA 5208

Qy        649 ATTGTTAATTTGTTAAATGAGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCA 708
              || ||  ||| ||  || || | | |  | |    |  | ||    |     | | |
Db       5209 TTTTTTTTTTTTTTTTTATTAGGAAGATTTTTTTTAATATAGTTAAATATTTTTTTTTTTA 5268

Qy        709 CTTAGTTTGTTTGATAGGCTCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCT 768
              |   |||  || |      || ||  |||   ||||   |    |  |
Db       5269 AGTTTAATGTAGGAGAAAAGTTTATTAAAATTTTATTTATGAATTTTTTGATATATCGTA 5328

Qy        769 TATGTCCCTTGCCTTCAACTAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAA 828
              | | |         || |  |   | ||    | |  | | ||  |  | |    |  |
Db       5329 TTTTTAAAGGTATTTTATTTTTTTTAAAATGTTTAGAGTTTTGTTTTTTATATTTAATAG 5388

Qy        829 ATATTACTACCAATGTTTGCTAGAATGAAAGCATCTAACACATTTCCCTCTAATTGTTTT 888
              | | |    | ||   |  ||| | |    ||    | ||     | |||||||
Db       5389 TTCGTTATTTAGAATTTATTTTTTATGTATGTTGTTAGATAAGTTTTTAATTTTTGTTTT 5448

```
Qy   889 TTTAAAATTCTGCTTGTATCACTAATTTCTGTTTTGCCTTGTCTTGCGTTATTCTTTTTT 948
              || |||| | |||| | | | | || || | | || |||||| || |||
Db   5449 GTTTTAATTTTAGTTGTTTTAATGTGATTTGATTGTATTATTTTTTTGTTATTTTTATTT 5508

Qy   949 GGTCGTGATATATTATCAATATGGATAAACCCTACATTTGCAACTGAAAATTATAAATTA 1008
              | | | | |||| ||||| | | | || || | | | |||
Db   5509 GTTATTTTAAGTTTATAGATATGTGTGGAGTTAATTTTATTTTATGGATTTATTTTTTTA 5568

Qy   1009 ATGCAAATTTTAGCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCATTTCTTG 1068
              | | || | || || | || ||| | || | | |||| || | | |
Db   5569 TTTTTGAATTATTATTTATTTTTTTTAATAGTGTAGTTTATTATAATTTTTGATGTTTGG 5628

Qy   1069 TTATTAGGAATTGGTAAATCTAAGCTTGTTGCAAATTT 1106
                 | | | || || | | || || |||
Db   5629 GGTTGTGTTTTCGGGAATGTTTTGGTTATTTTTGTTTT 5666

RESULT 8
US-10-240-708-63
; Sequence 63, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 63
; LENGTH: 5562
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-63

Query Match 4.3%; Score 55; DB 10; Length 5562;
Best Local Similarity 47.3%; Pred. No. 5.5;
Matches 239; Conservative 0; Mismatches 255; Indels 11; Gaps 2;

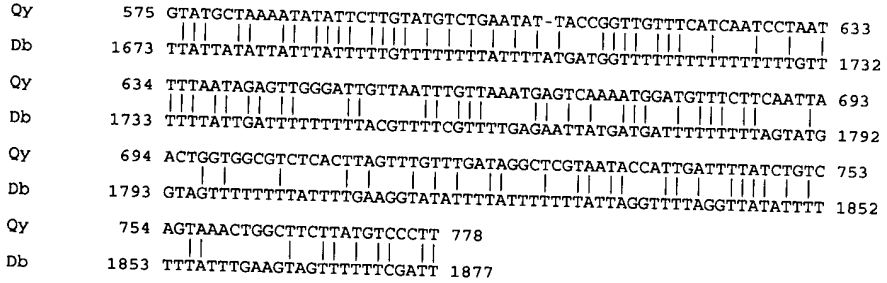
```
Qy        275 TTTCCTATAATTTTTCAATTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATAT 334
              | |  | | | |  ||  ||| ||||| |    |||  ||||   ||  |  |||| || |
Db       1383 TATTATTTTAATATTTTATTATTATTTTTTTTTTTTCGGTTAGGAATGATTTTTTTTTTG 1442

Qy        335 TAAGTATTTGTTTCTTCGATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTA 394
                  |||||  | | ||      ||  |   | |  ||  ||  ||  |  ||      ||||||
Db       1443 ATTATATTTATATTTTTTTATTTTAGTTTAGTTTTTTTTTTTTTTTTATTTATTTTTTATTA 1502

Qy        395 TTTGTGGTTTTTTTCAGGAAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAA 454
              |||  |  |||||  | ||  || |  |  |||  |||| |  |  |      |||||
Db       1503 TTTTTATTTTTTGTTAGTTAGATTTTTTTTTTTTTTATTATTT----------TTTTGTTT 1552

Qy        455 AATTTAATGTATCATGTTTTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAA 514
                | ||  |||  |  | ||  ||||     |  ||  |  |    | |    ||  |||  |
Db       1553 TTTATAGTGTTTTTTTTTATTTTTTTTTTGTTTAAGAAAGTTTATTAATTTTTATTTTTT 1612

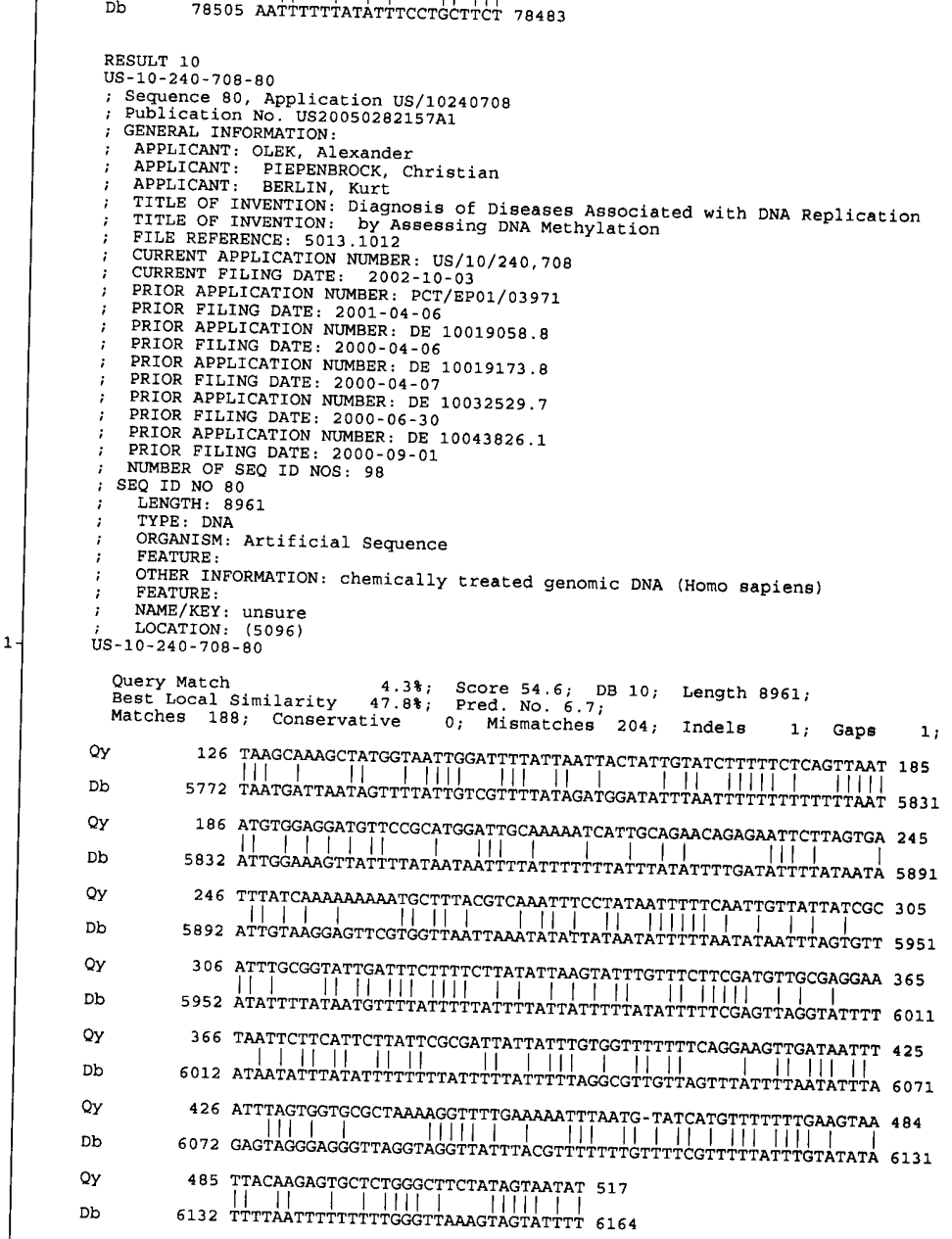
Qy        515 TATATGGCATTTACGGAAATGCACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAG 574
              |          |||         | |   |  ||||||     |||  |   | |  || |
Db       1613 TTATATTTTTTTTTTTTTTAGGATAGTGTTATTTTTAGAATTTTGTGATAATAATGATTT 1672
```



RESULT 9
US-10-995-561-13303/c
; Sequence 13303, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13303
; LENGTH: 201990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(201990)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13303

Query Match 4.3%; Score 55; DB 10; Length 201990;
Best Local Similarity 47.0%; Pred. No. 8.1;
Matches 208; Conservative 0; Mismatches 230; Indels 5; Gaps 1;





RESULT 10
US-10-240-708-80
; Sequence 80, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 80
; LENGTH: 8961
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (5096)
US-10-240-708-80

Query Match 4.3%; Score 54.6; DB 10; Length 8961;
Best Local Similarity 47.8%; Pred. No. 6.7;
Matches 188; Conservative 0; Mismatches 204; Indels 1; Gaps 1;

RESULT 11
US-10-240-708-35
; Sequence 35, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 35
; LENGTH: 9347
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-35

Query Match 4.3%; Score 54.2; DB 10; Length 9347;
Best Local Similarity 45.1%; Pred. No. 7.9;
Matches 243; Conservative 0; Mismatches 293; Indels 3; Gaps 1;

```
Qy       127 AAGCAAAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATA 186
Db      7998 AAAAAAGGGGGGGGGTTTTGTTATTTTTAAATATTTTTTTTTTAAATTTATTTTAGTAAG 8057

Qy       187 TGTGGAGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGAGAATTCTTAGTGAT 246
Db      8058 TTTTTTTGTTGTTATTATTTTATTAAAGTTGTTTTATTTAAGGTTATTAGTGATTTTTAT 8117

Qy       247 TTATCAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCA 306
Db      8118 TATGTTAAAATTAGTGATGAATTTTTAGTTTTTATTTTGAATTGAATTAGTATCGGTATT 8177

Qy       307 TTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAAT 366
Db      8178 TGATATTGTTTGTTATTTTTTTTTTTTTTTAAATATATTTTTTTTTTAGTTTTTTAGATAT 8237

Qy       367 AATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTA 426
Db      8238 TATTAGTTTTTTTTTTTTTTTTTTTTGAGATAGGGTTTTGTTTTGTTGTTTAGGTTGGAG 8297

Qy       427 TTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATT 486
Db      8298 TATAATGGGTAAATTATAGTTTATTGTAGTTTTAATTTTTTTGGTTTAAGTAATTTTTTT 8357

Qy       487 ACAAGAGTGCT---CTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTA 543
Db      8358 ATTTTAGTTTTTAAGTAGTTGTAATTATAGGTATATGTTATTAAGTTTATTTAATTGTTG 8417

Qy       544 TATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTG 603
Db      8418 TATTTTTTATAGAAACGGTTTTATTTTGTTGTTTAAGTTGGTTTTAAATTTTTGGGTTTA 8477

Qy       604 AATATTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTT 662
Db      8478 AATAGTTTGTTTATTTTAATTTTATAAAATATTGGATTATAGGCGTGAGTTATTGTATT 8536
```

RESULT 12
US-11-011-332A-155
; Sequence 155, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
; APPLICANT: Foekens, John
; APPLICANT: Harbeck, Nadia
; APPLICANT: Koenig, Thomas
; APPLICANT: Maier, Sabine
; APPLICANT: Martens, John
; APPLICANT: Model, Fabian
; APPLICANT: Nimmrich, Inko
; APPLICANT: Rujan, Tamas
; APPLICANT: Schmitt, Manfred
; APPLICANT: Lesche, Ralf
; APPLICANT: Dietrich, Dimo
; APPLICANT: Mueller, Volkmar
; APPLICANT: Kluth, Antje
; APPLICANT: Schwope, Ina
; APPLICANT: Hartmann, Oliver
; APPLICANT: Adorjan, Peter
; TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SUR
; TITLE OF INVENTION: BREAST CELL PROLIFERATIVE DISORDER PATIENTS
; FILE REFERENCE: 47675-99
; CURRENT APPLICATION NUMBER: US/11/011,332A
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 10/517,741
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: PCT/EP2003/010881
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: DE 10245779.4
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: DE 10300096.8
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: DE 10317955.0
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: PCT/EP2004/014170
; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: EP 03090432.0
; PRIOR FILING DATE: 2003-12-11
; PRIOR APPLICATION NUMBER: EP 04090041.7
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: EP 04090380.9
; PRIOR FILING DATE: 2004-09-30
; PRIOR APPLICATION NUMBER: EP 04090127.4
; PRIOR FILING DATE: 2004-04-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 155
; LENGTH: 28536
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-155

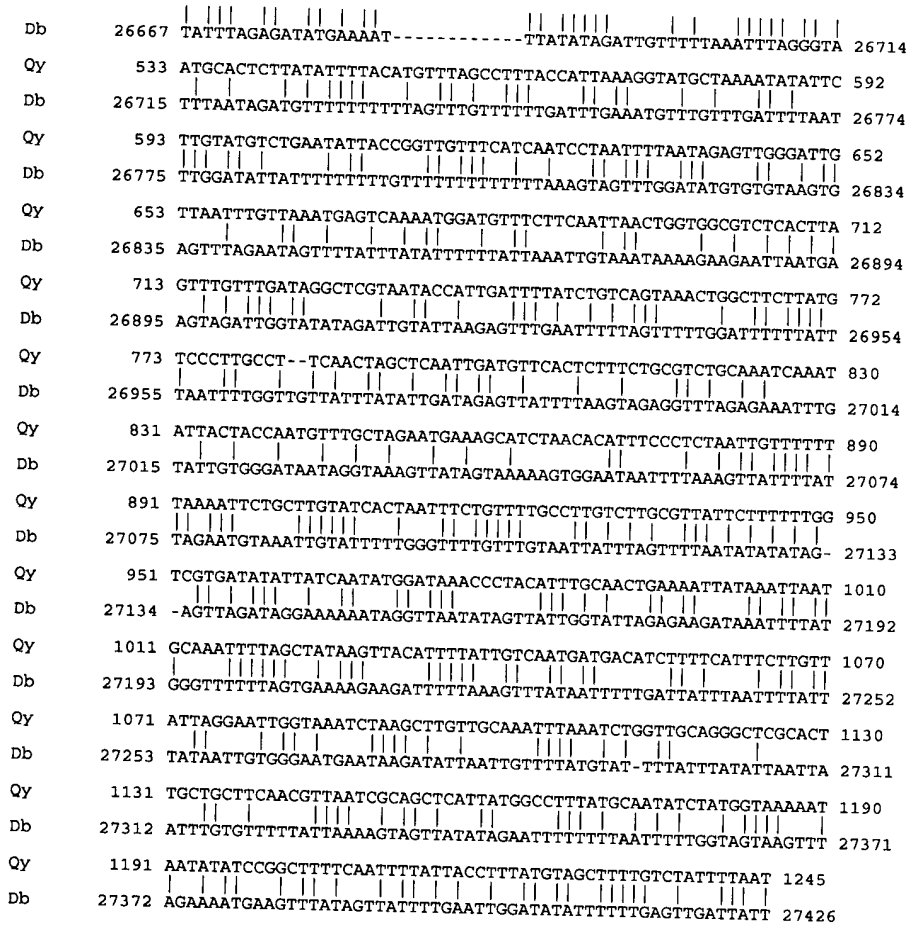
Query Match 4.3%; Score 54.2; DB 17; Length 28536;
Best Local Similarity 44.8%; Pred. No. 8.9;
Matches 428; Conservative 0; Mismatches 508; Indels 19; Gaps 5;

```
Qy       293 TTGTTATTATCGCATTTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCG 352
Db     26489 TTTATTTTTTAGTGGTTTTTTTTTTGTAGTAGTTTTTTATTTGAATTTTTTTTTGTTTT 26548

Qy       353 ATGTTGCGAGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGG 412
Db     26549 TTTGTGGTAGGGTTTGTATATTGATTTTTTTGATTTTTGGTATATTTGGGTTTTTT--GA 26606

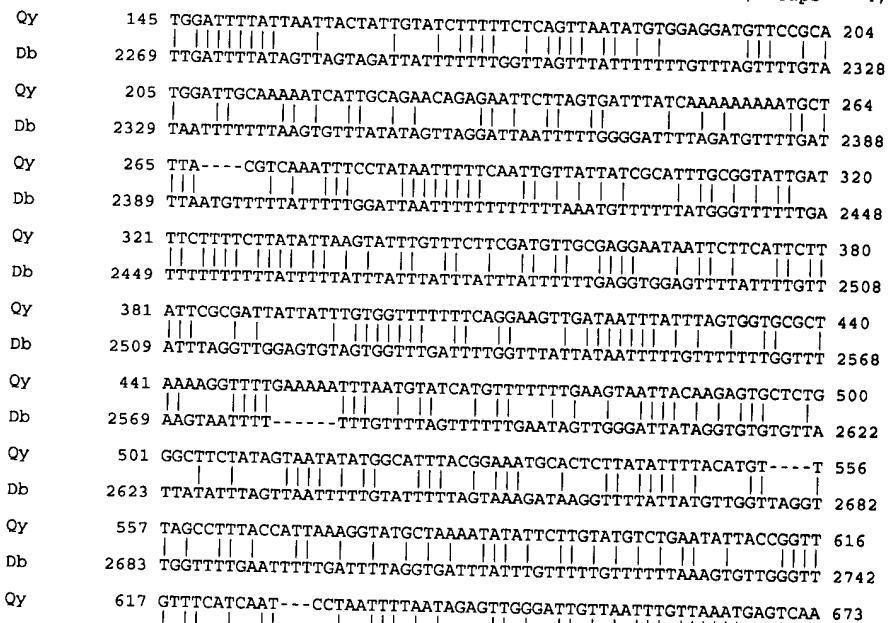
Qy       413 AAGTTGATAATTTATTTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTT 472
Db     26607 AATTTTTTAATTTTTTTAGATTTGAGGATGGTAGGTTTTATTTTTTTTTATTGTGTGTATA 26666

Qy       473 TTTTTGAAGTAATTACAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAA 532
```



RESULT 13
US-11-011-332A-91
; Sequence 91, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
; APPLICANT: Foekens, John
; APPLICANT: Harbeck, Nadia
; APPLICANT: Koenig, Thomas
; APPLICANT: Maier, Sabine
; APPLICANT: Martens, John
; APPLICANT: Model, Fabian
; APPLICANT: Nimmrich, Inko
; APPLICANT: Rujan, Tamas
; APPLICANT: Schmitt, Manfred
; APPLICANT: Lesche, Ralf
; APPLICANT: Dietrich, Dimo
; APPLICANT: Mueller, Volkmar
; APPLICANT: Kluth, Antje
; APPLICANT: Schwope, Ina
; APPLICANT: Hartmann, Oliver
; APPLICANT: Adorjan, Peter
; TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SUR
; TITLE OF INVENTION: BREAST CELL PROLIFERATIVE DISORDER PATIENTS
; FILE REFERENCE: 47675-99
; CURRENT APPLICATION NUMBER: US/11/011,332A
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 10/517,741
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: PCT/EP2003/010881
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: DE 10245779.4
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: DE 10300096.8
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: DE 10317955.0
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: PCT/EP2004/014170
; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: EP 03090432.0
; PRIOR FILING DATE: 2003-12-11
; PRIOR APPLICATION NUMBER: EP 04090041.7
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: EP 04090380.9
; PRIOR FILING DATE: 2004-09-30
; PRIOR APPLICATION NUMBER: EP 04090127.4
; PRIOR FILING DATE: 2004-04-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 91
; LENGTH: 6499
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-91

Query Match 4.2%; Score 53.4; DB 17; Length 6499;
Best Local Similarity 48.7%; Pred. No. 10;
Matches 273; Conservative 0; Mismatches 271; Indels 17; Gaps 4;



```
Db       2743 GATTTTTTTATTTGGTTTGTTTGATTTAGATTTGTTTTTTAATTTTATTAAATAGTTAGG 2802

Qy        674 AATGGATGTTTCTTCAATTAA 694
                | ||| ||| ||    |||
Db       2803 TGTAGATTTTTTTTTTGGTAA 2823


RESULT 14
US-10-793-626-3976/c
; Sequence 3976, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3976
;   LENGTH: 4069
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3976

  Query Match             4.2%;  Score 52.8;  DB 10;  Length 4069;
  Best Local Similarity   45.0%;  Pred. No. 12;
  Matches  242;  Conservative    0;  Mismatches  292;  Indels    4;  Gaps    1;

Qy        132 AAGCTATGGTAATTGGATTTTATTAATTACTATTGTATCTTTTTCTCAGTTAATATGTGG 191
              || |  || ||||| ||||| || || |  ||||     ||| |||     |     | |
Db       2470 AAACCCTGTTAATTCGATTTCATGAAATCATATTAATGATTTATCTGTTTAGTATTATAT 2411

Qy        192 AGGATGTTCCGCATGGATTGCAAAAATCATTGCAGAACAGA----GAATTCTTAGTGATT 247
                |  ||    ||    |  | |||   |   |||  |||     |  | ||  |   |
Db       2410 GTTACTTTGTTGATTTCATTAATAAAAATTATAAGACTAGAACTGTAGGTTTTTATTCAT 2351

Qy        248 TATCAAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCAT 307
              ||   |          | |||||    ||| || |    | |||   || ||  ||    |
Db       2350 TAGGGATTGTTTGGGTTTTACAAACAATCTCTTTGTCTATTTTTTTTATTCATACTAAAC 2291

Qy        308 TTGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATA 367
              | |  | |   |||| |  ||    |   | | | |||        ||   |     |
Db       2290 AAGTGCCACTAGGTTCTATATTTGACGTTTTTTTCTCTTTAACTTGGATTATTATATCAA 2231

Qy        368 ATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTAT 427
              |||  |    | ||       |||||    |       |||||  |        |  ||||||
Db       2230 TTTCGTTGATATTGAATTTGATTAAAGTAATGAATTTTTCTGTTTTTTTCCTTAATTTGA 2171

Qy        428 TTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTA 487
              || |    |  ||||   ||||  | |  ||| |      |  || |          ||||
Db       2170 TTGGATTTATCCTAATGAGTTTAAATACTTTTCAGCCTGAACATTATCAAACGCAAATTC 2111

Qy        488 CAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATT 547
              |  | |          |  |  | || ||  ||     |   | |  |||| | |||  ||
Db       2110 AACAAATTGCAGTAATTAATGAATTATTACTTGTTCATATAGCACTTGCAGTATTAAGTT 2051

Qy        548 TTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATA 607
              | ||| ||| ||  |  |  ||  |  |     ||   || ||| ||  || ||  |||
Db       2050 ATGCATTTTTTGCAATCGCATTTGTAAATTCATTACTCTACATTATTCAATATCGAAATT 1991

Qy        608 TTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAA 665
              | |  |          || |||  ||||    |||| |||| | ||||  |   ||| |
Db       1990 TAAAGGAGAAAAATTTCGATCAAAATTACTTTAGAATTGGTAGTGTTGCTACTTTAGA 1933


RESULT 15
US-10-240-708-73
; Sequence 73, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/240,708
;  CURRENT FILING DATE:  2002-10-03
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 73
;   LENGTH: 5152
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-73

  Query Match             4.1%;  Score 51.4;  DB 10;  Length 5152;
  Best Local Similarity   42.8%;  Pred. No. 21;
  Matches  420;  Conservative    0;  Mismatches  556;  Indels    5;  Gaps    3;

Qy        249 ATCAAAAAAAAAATGCTTTACGTCAAATTTCCTATAATTTTTCAATTGTTATTATCGCATT 308
              || ||||| | ||| ||||||   ||  |    ||   ||    || ||||  |    ||
Db       1707 ATTAAAAATAGATGTTTTACGGATAAAGTATATTATAATTAATTTTTTTATGTTAATTTT 1766

Qy        309 TGCGGTATTGATTTCTTTTCTTATATTAAGTATTTGTTTCTTCGATGTTGCGAGGAATA- 367
              |       |  ||| ||   | || ||| | |  |||    |  ||       |  ||||
Db       1767 TTATAATTAATTTTTTTAGTTAATTTTATGAAAATGTAGTATTTATAGAATTAAAAATAT 1826

Qy        368 ATTCTTCATTCTTATTCGCGATTATTATTTGTGGTTTTTTTCAGGAAGTTGATAATTTAT 427
              || || | |   |  |  |||| ||| || |    |||||          ||| ||  |
Db       1827 TTTTTTTAATGATTATTTTGATTTTTAATTTTTTAGTTTTTTGATTTTAAGATTATAAAA 1886

Qy        428 TTAGTGGTGCGCTAAAAGGTTTTGAAAAATTTAATGTATCATGTTTTTTTGAAGTAATTA 487
              ||   | ||   |  |   | | |  | |||      ||   |||||  |  || || | |
Db       1887 TTTAAGTTGTTTTTTATTTATATTATTATTTTTTGGTGATTTGGTTTTATAAAATTATTAA 1946

Qy        488 CAAGAGTGCTCTGGGCTTCTATAGTAATATATGGCATTTACGGAAATGCACTCTTATATT 547
                | |       ||  |   || || ||    ||||   ||        |  | |
Db       1947 GGGATTGGGTGAAAAGTTTAAGTATATTAAATTTATTTTAAATTAAATTTTAATGTTTTG 2006

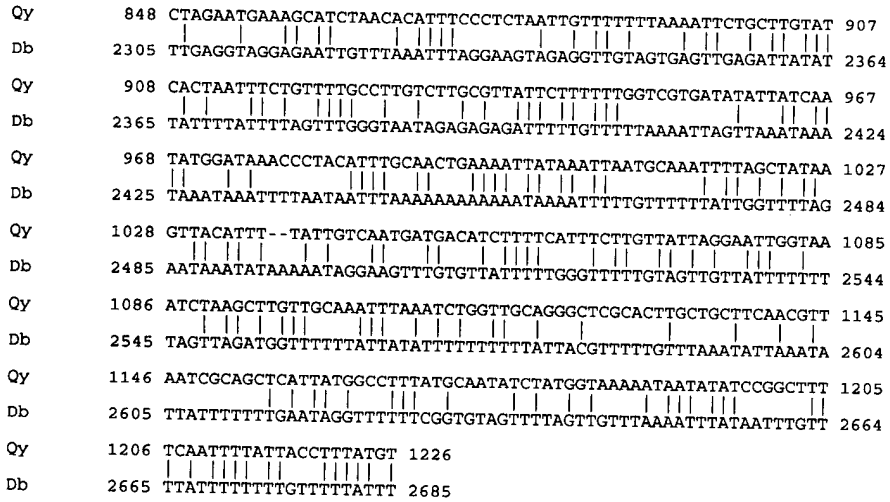
Qy        548 TTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAAATATATTCTTGTATGTCTGAATA 607
                |  | |   ||       | | | |     |||   |||    ||     |   ||
Db       2007 AAGGTTTATAAAAGTTATATTTGAGAAGATATAAAAATATTATAGGAGTGAAGGTTTGTA 2066

Qy        608 TTACCGGTTGTTTCATCAATCCTAATTTTAATAGAGTTGGGATTGTTAATTTGTTAAATG 667
              |||  |     || || |||  |||||   ||||  ||    |   || || |   |  |
Db       2067 TTATTGTAGTATTTATAAATTTTAATTAATATAGTATTTTTAAATTTTATATTATTTAAG 2126

Qy        668 AGTCAAAATGGATGTTTCTTCAATTAACTGGTGGCGTCTCACTTAGTTTGTTTGATAGGC 727
              | |  |||| |    | |    ||   ||  |  || |   ||  |   |||  ||
Db       2127 ATTTTAAATTTAATAATTTATGGTTGGATGAGGTGGTTTATGTTTATAATTTTAGTATTT 2186

Qy        728 TCGTAATACCATTGATTTTATCTGTCAGTAAACTGGCTTCTTATGTCCCTTGCCTTCAAC 787
              | | |   | |  |  | |  |||  |   | |    |   | |   ||  || ||
Db       2187 TGGGAGGTCGAGGTAGGTGAATTGTTTGAGTAATTTGGTAAAGTTTTATTTTTATTAAAA 2246

Qy        788 TAGCTCAATTGATGTTCACTCTTTCTGCGTCTGCAAATCAAATATTACTACCAATGTTTG 847
                    ||    |||   | | |  |||| ||    ||   | ||| |     |  |
Db       2247 ATATAAAAAATTAGTTGGGTATATTGGCGTGTGTTTGTCG--TTTTAGTTATTCGGGAGG 2304
```



Search completed: May 11, 2006, 08:30:37
Job time : 656.207 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:45:22 ; Search time 3822.17 Seconds
(without alignments)
15660.244 Million cell updates/sec

Title: US-09-423-093B-1_COPY_9901_10953
Perfect score: 1053
Sequence: 1 atgtctattgatttacttttt.....gatttgcaaaaattagataa 1053

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1053	100.0	14516	1	AF078736	AF078736 Escherich
	2	1053	100.0	14516	6	BD085554	BD085554 Nucleic a
	3	1053	100.0	14520	6	AX029261	AX029261 Sequence
	4	685.6	65.1	13650	1	AF285969	AF285969 Salmonell
	5	133.8	12.7	1328	15	AJ840631	AJ840631 Arabidops
	6	121.4	11.5	282289	14	AC152242	AC152242 Bos tauru
	7	119.4	11.3	66993	14	AC138074	AC138074 Homo sapi
	8	118	11.2	1142	15	AJ840525	AJ840525 Arabidops
	9	116.8	11.1	93791	14	AC138073	AC138073 Homo sapi
	10	116.6	11.1	15387	6	AX345086	AX345086 Sequence
	11	116.6	11.1	170627	14	AC125567	AC125567 Rattus no
c	12	113.4	10.8	154625	14	CR936360	CR936360 Homo sapi
	13	113.2	10.8	7442	6	AX251441	AX251441 Sequence
c	14	111.6	10.6	67970	2	PFMAL1P3	AL031746 Plasmodiu
c	15	110.6	10.5	175544	14	AC117342	AC117342 Rattus no
	16	109.2	10.4	223697	14	AC151103	AC151103 Bos tauru
	17	107.6	10.2	7814	6	AX251284	AX251284 Sequence
	18	107.4	10.2	7597	6	AX345915	AX345915 Sequence
	19	107	10.2	6668	6	AX346599	AX346599 Sequence
	20	106	10.1	157141	2	AC016445	AC016445 Drosophil
	21	105.6	10.0	1197	15	AJ840663	AJ840663 Arabidops
c	22	105.6	10.0	6644	6	E23356	E23356 Virus vecto
c	23	105.6	10.0	7372	6	E23357	E23357 Virus vecto
c	24	105.6	10.0	7797	6	E23355	E23355 Virus vecto
c	25	105.6	10.0	7996	6	E23359	E23359 Virus vecto
	26	105.6	10.0	9539	6	AX277889	AX277889 Sequence
	27	105.6	10.0	9539	6	AX323566	AX323566 Sequence
c	28	105.6	10.0	289973	14	AC135678	AC135678 Rattus no
	29	104.8	10.0	286208	14	AC117140	AC117140 Rattus no
c	30	103.2	9.8	84563	5	AC096885	AC096885 Danio rer
c	31	103.2	9.8	200187	14	CR925711	CR925711 Danio rer
c	32	103	9.8	162115	5	BX569788	BX569788 Zebrafish
c	33	102.8	9.8	170164	9	AC110382	AC110382 Mus muscu
	34	102.2	9.7	6292	6	AX251493	AX251493 Sequence
	35	102	9.7	349980	6	AX344563	AX344563 Sequence
	36	101.8	9.7	13511	6	AX345183	AX345183 Sequence
	37	101.6	9.6	5493	6	CQ807241	CQ807241 Sequence
c	38	101.6	9.6	212999	14	AC151201	AC151201 Bos tauru
	39	101	9.6	54345	2	AC084152	AC084152 Caenorhab
c	40	100.8	9.6	83435	5	BX640469	BX640469 Zebrafish
	41	100.8	9.6	153759	14	CR848039	CR848039 Danio rer
c	42	100.8	9.6	160405	5	BX510999	BX510999 Zebrafish
	43	100.6	9.6	47108	6	AX344507	AX344507 Sequence
	44	100.6	9.6	145925	5	BX842571	BX842571 Zebrafish
	45	100.4	9.5	9539	6	AX277888	AX277888 Sequence

ALIGNMENTS

RESULT 1
AF078736
LOCUS AF078736 14516 bp DNA linear BCT 03-AUG-1999
DEFINITION Escherichia coli O111 O antigen gene cluster, partial sequence.
ACCESSION AF078736 U13629
VERSION AF078736.1 GI:5678713
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 14516)
AUTHORS Bastin,D.A. and Reeves,P.R.
TITLE Sequence and analysis of the O antigen gene (rfb) cluster of Escherichia coli O111
JOURNAL Gene 164 (1), 17-23 (1995)
PUBMED 7590310
REFERENCE 2 (bases 3021 to 9981)
AUTHORS Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C., Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
TITLE Bacterial polysaccharide synthesis and gene nomenclature
JOURNAL Trends Microbiol. 4 (12), 495-503 (1996)
PUBMED 9004408
REFERENCE 3 (bases 1 to 14516)
AUTHORS Wang,L., Curd,H., Qu,W. and Reeves,P.R.
TITLE Sequencing of Escherichia coli O111 O-antigen gene cluster and identification of O111-specific genes
JOURNAL J. Clin. Microbiol. 36 (11), 3182-3187 (1998)
PUBMED 9774562
REFERENCE 4 (bases 3021 to 9981)
AUTHORS Bastin,D.A.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-1997) Microbiology, University of Sydney, Maze Crescent, Sydney, NSW 2006, Australia
REFERENCE 5 (bases 1 to 14516)
AUTHORS Wang,L., Curd,H., Qu,W., Bastin,D.A. and Reeves,P.R.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-1998) Microbiology, University of Sydney, Maze Crescent, Sydney, NSW 2006, Australia
COMMENT On Aug 3, 1999 this sequence version replaced gi:1155016.
FEATURES Location/Qualifiers

```
source          1. .14516
                /organism="Escherichia coli"
                /mol_type="genomic DNA"
                /strain="Stoke W; M92"
                /serotype="O111"
                /db_xref="taxon:562"
gene            <1. .273
                /gene="galF"
CDS             <1. .273
                /gene="galF"
                /codon_start=1
                /transl_table=11
                /product="putative glucose-1-phosphate
                uridylyltransferase"
                /protein_id="AAD46727.1"
                /db_xref="GI:5678714"
                /translation="DLMAVGRYVLSADIWAELEKTAPGAWGRIQLTDAIAELAKKQSV
                DAMLMTGDSYDCGKKMGYMQAFVKYGLRNLKEGAKFRKSIKKLLSE"
gene            739. .1932
                /gene="wbdH"
CDS             739. .1932
                /gene="wbdH"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase"
                /protein_id="AAD46728.1"
                /db_xref="GI:5678715"
                /translation="MLLCCIHINVYYLLLECDMKKIVIIGNVASMMLRFRKELIMNLV
                RQGDNVYCLANDFSTEDLKVLSSWGVKGVKFSLNSKGINPFKDIIAVYELKKILKDIS
                PDIVFSYFVKPVIFGTIASKLSKVPRIVGMIEGLGNAFTYYKGKQTTKTKMIKWIQIL
                LYKLALPMLDDLILLLNHDDKKDLIDQYNIKAKVTVLGGIGLDLNEFSYKEPPKEKITF
                IFIARLLREKGIFEFIEAAKFVKTTYPSSEFVILGGFESNNPFSLQKNEIESLRKEHD
                LIYPGHVENVQDWLEKSSVFVLPTSYREGVPRVIQEAMAIGRPVITTNVPGCRDIIND
                GVNGFLIPPFEINLLAEKMKYFIENKDKVLEMGLAGRKFAEKNFDAFEKNNRLASIIK
                SNNDF"
gene            1983. .3101
                /gene="gmd"
CDS             1983. .3101
                /gene="gmd"
                /codon_start=1
                /transl_table=11
                /product="GDP-D-mannose dehydratase"
                /protein_id="AAD46729.1"
                /db_xref="GI:5678716"
                /translation="MNKVALITGITGQDGSYLAELLLEKGYEVHGIKRRASSFNTERV
                DHIYQDSHLANPKLFLHYGDLTDTSNLTRILKEVQPDEVYNLGAMSHVAVSFESPEYT
                ADVDAIGTLRLLEAIRILGLEKKTKFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
                KLYAYWITVNYRESYGMFACNGILFNHESPRRGETFVTRKITRGIANIAQGLDKCLYL
                GNMDSLRDWGHAKDYVKMQWMMLQQETPEDFVIATGIQYSVREFVTMAAEQVGIELAF
                EGEGVNEKGVVVSVNGTDAKAVNPGDVIISVDPRYFRPAEVETLLGDPTNAHKKLGWS
                PEITLREMVKEMVSSDLAIAKKNVLLKANNIATNIPQE"
gene            3174. .3623
                /gene="wbdI"
CDS             3174. .3623
                /gene="wbdI"
                /note="previously known as orf3.4"
                /codon_start=1
                /transl_table=11
                /product="WbdI"
                /protein_id="AAC44881.1"
                /db_xref="GI:1155017"
                /translation="MFITSDKFREIIKLVPLVSIDLLIENENGEYLFGLRNNRPAKNY
                FFVPGGRIRKNESIKNAFKRISSMELGKEYGISGSVFNGVWEHFYDDGFFSEGEATHY
                IVLCYTLKVLKSELNLPDDQHREYLWLTKHQINAKQDVHNYSKNYFL"
gene            3657. .5108
                /gene="manC"
CDS             3657. .5108
                /gene="manC"
                /EC_number="2.7.7.22"
                /note="previously known as RfbM"
                /codon_start=1
                /transl_table=11
                /product="mannose-1-phosphate guanylyltransferase"
                /protein_id="AAC44882.1"
                /db_xref="GI:1155018"
                /translation="MSQCLYPVIIAGGTGSRLWPLSRVLYPKQFLNLVGDSTMLQTTI
                TRLDGIECENPIVICNEDHRFIVAEQLRQIGKLTKNIILEPKGRNTAPAIALAAFIAQ
                KNNPNDDPLLLVLAADHSINNEKAFRESIIKAMPYATSGKLVTFGIIPDTANTGYGYI
                KRSSSADPNKEFPAYNVAEFVEKPDVKTAQEYISSGNYYWNSGMFLFRASKYLDELRK
                FRPDIYHSCECATATANIDMDFVRINEAEFINCPEESIDYAVMEKTKDAVVLPIDIGW
                NDVGSWSSLWDISQKDCHGNVCHGDVLNHDGENSFIYSESSLVATVGVSNLVIVQTKD
                AVLVADRDKVQNVKNIVDDLKKRKRAEYYMHRAVFRPWGKFDAIDQGDRYRVKKIIVK
                PGEGLDLRMHHHRAEHWIVVSGTAKVSLGSEVKLLVSNESIYIPQGAKYSLENPGVIP
                LHLIEVSSGDYLESDDIVRFTDRYNSKQFLKRD"
gene            5114. .6484
                /gene="manB"
CDS             5114. .6484
                /gene="manB"
                /EC_number="5.4.2.8"
                /note="previously known as RfbK"
                /codon_start=1
                /transl_table=11
                /product="phosphomannomutase"
                /protein_id="AAC44883.1"
                /db_xref="GI:1155019"
                /translation="MNKITCFKAYDIRGRLGAELNDEIAYRIGRAYGEFFKPQTVVVG
                GDARLTSESLKKSLSNGLCDAGVNVLDLGMCGTEEIYFSTWYLGIDGGIEVTASHNPI
                DYNGMKLVTKGARPISSDTGLKDIQQLVESNNFEELNLEKKGNITKYSTRDAYINHLM
                GYANLQKIKKIKIVVNSGNGAAGPVIDAIEECFLRNNIPIQFVKINNTPDGNFPHGIP
                NPLLPECREDTSSAVIRHSADFGIAFDGDFDRCFFFDENGQFIEGYYIVGLLAEVFLG
                KYPNAKIIHDPRLIWNTIDIVESHGGIPIMTKTGHAYIKQRMREEDAVYGGEMSAHHY
                FKDFAYCDSGMIPWILICELLSLTNKKLGELVCGCINDWPASGEINCTLDNPQNEIDK
                LFNRYKDSALAVDYTDGLTMEFSDWRFNVRCSNTEPVVRLNVESRNNAILMQEKTEEI
                LNFISK"
gene            6521. .7444
                /gene="wbdJ"
CDS             6521. .7444
                /gene="wbdJ"
                /note="previously known as orf6.7"
                /codon_start=1
                /transl_table=11
                /product="WbdJ"
                /protein_id="AAC44884.1"
                /db_xref="GI:1155020"
                /translation="MKVLLTGSTGMVGKNILEHDSASKYNILTPTSSDLNLLDKNEIE
                KFMLINMPDCIIHAAGLVGGIHANISRPFDFLEKNLQMGLNLVSVAKKLGIKKVLNLG
                SSCMYPKNFEEAIPEKALLTGELEETNEGYAIAKIAVAKACEYISRENSNYFYKTIIP
                CNLYGKYDKFDDNSSHMIPAVIKKIHHAKINNVPEIEIWGDGNSRREFMYAEDLADLI
                FYVIPKIEFMPNMVNAGLGYDYSINDYYKIIAEEIGYTGSFSHDLTKPTGMKRKLVDI
                SLLNKIGWSSHFELRDGIRKTYNYYLENQNK"
gene            7441. .8607
                /gene="wbdK"
CDS             7441. .8607
                /gene="wbdK"
                /note="previously known as orf7.7"
                /codon_start=1
                /transl_table=11
                /product="WbdK"
                /protein_id="AAC44885.1"
                /db_xref="GI:1155021"
                /translation="MITYPLASNTWDEYEYAAIQSVIDSKMFTMGKKVELYEKNFADL
                FGSKYAVMVSSGSTANLLMIAALFFTNKPKLKRGDEIIVPAVSWSTTYYPLQQYGLKV
                KFVDINKETLNIDIDSLKNAISDKTKAILTVNLLGNPNDFAKINEIINNRDIILLEDN
                CESMGAVFQNKQAGTFGVMGTFSSFYSHHIATMEGGCVVTDDEELYHVLLCLRAHGWT
                RNLPKENMVTGTKSDDIFEESFKFVLPGYNVRPLEMSGAIGIEQLKKLPGFISTRRSN
                AQYFVDKFKDHPFLDIQKEVGESSWFGFSFVIKEGAAIERKSLVNNLISAGIECRPIV
                TGNFLKNERVLSYFDYSVHDTVANAEYIDKNGFFVGNHQIPLFNEIDYLRKVLK"
gene            8646. .9911
                /gene="wzx"
CDS             8646. .9911
                /gene="wzx"
                /note="integral membrane protein; previously known as
                orf8.9"
                /codon_start=1
                /transl_table=11
                /product="putative O unit flippase"
```

```
                /protein_id="AAC44886.1"
                /db_xref="GI:1155022"
                /translation="MVLTVKKILAFGYSKVLPPVIEQFVNPICIFIITPLILNHLGKQ
                SYGNWILLITIVSFSQLICGGCSAWIAKIIAEQRILSDLSKKNALRQISYNFSIVIIA
                FAVLISFLILSICFFDVARNNSSFLFAIIICGFFQEVDNLFSGALKGFEKFNVSCFFE
                VITRVLWASIVIYGIYGNALLYFTCLAFTIKGMLKYILVCLNITGCFINPNFNRVGIV
                NLLNESKWMFLQLTGGVSLSLFDRLVIPLILSVSKLASYVPCLQLAQLMFTLSASANQ
                ILLPMFARMKASNTFPSNCFFKILLVSLISVLPCLALFFFGRDILSIWINPTFATENY
                KLMQILAISYILLSMMTSFHFLLLGIGKSKLVANLNLVAGLALAASTLIAAHYGLYAI
                SMVKIIYPAFQFYYLYVAFVYFNRAKNVY"
     gene       9901. .10953
                /gene="wzy"
     CDS        9901. .10953
                /gene="wzy"
                /codon_start=1

  Query Match             100.0%;  Score 1053;  DB 1;  Length 14516;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-129;
  Matches 1053;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9901 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 9960

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9961 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 10020

Qy        121 CTTTTATGCTTGCTCTTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10021 CTTTTATGCTTGCTCTTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 10080

Qy        181 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10081 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 10140

Qy        241 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10141 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 10200

Qy        301 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10201 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTG 10260

Qy        361 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10261 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 10320

Qy        421 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10321 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 10380

Qy        481 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10381 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 10440

Qy        541 AAAAAGATTCCATACTTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10441 AAAAAGATTCCATACTTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT 10500

Qy        601 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10501 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT 10560

Qy        661 GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10561 GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT 10620

Qy        721 ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10621 ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA 10680

Qy        781 TCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10681 TCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTG 10740

Qy        841 ATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10741 ATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTA 10800

Qy        901 TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10801 TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTG 10860

Qy        961 TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10861 TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT 10920

Qy       1021 GGAGTTGTTTATGGATTTGCAAAAATTAGATAA 1053
              |||||||||||||||||||||||||||||||||
Db      10921 GGAGTTGTTTATGGATTTGCAAAAATTAGATAA 10953


RESULT 2
BD085554
LOCUS       BD085554               14516 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085554
VERSION     BD085554.1  GI:22631164
KEYWORDS    JP 2001522248-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 14516)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 1 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Escherichia coli
            PN   JP 2001522248-A/1
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14516
            FT                   /organism='Escherichia coli'.
FEATURES             Location/Qualifiers
     source          1. .14516
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 1053;  DB 6;  Length 14516;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-129;
  Matches 1053;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9901 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 9960

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9961 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 10020

Qy        121 CTTTTATGCTTGCTCTTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10021 CTTTTATGCTTGCTCTTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 10080

Qy        181 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      10081 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 10140

Qy        241 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10141 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 10200

Qy        301 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTTATATTGCTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10201 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTTATATTGCTG 10260

Qy        361 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10261 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 10320

Qy        421 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10321 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 10380

Qy        481 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10381 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 10440

Qy        541 AAAAAGATTCCATACTTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10441 AAAAAGATTCCATACTTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT 10500

Qy        601 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10501 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT 10560

Qy        661 GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10561 GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT 10620

Qy        721 ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10621 ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA 10680

Qy        781 TCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10681 TCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTG 10740

Qy        841 ATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10741 ATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTA 10800

Qy        901 TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10801 TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTG 10860

Qy        961 TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10861 TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT 10920

Qy       1021 GGAGTTGTTTATGGATTTGCAAAAATTAGATAA 1053
              |||||||||||||||||||||||||||||||||
Db      10921 GGAGTTGTTTATGGATTTGCAAAAATTAGATAA 10953
```

```
RESULT 3
AX029261
LOCUS       AX029261               14520 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 1 from Patent WO9850531.
ACCESSION   AX029261
VERSION     AX029261.1  GI:10190112
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 1 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .14520
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
ORIGIN
```

```
  Query Match           100.0%;  Score 1053;  DB 6;  Length 14520;
  Best Local Similarity  100.0%;  Pred. No. 7.6e-129;
  Matches 1053;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9901 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 9960

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9961 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 10020

Qy        121 CTTTTATGCTTGCTCTTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10021 CTTTTATGCTTGCTCTTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 10080

Qy        181 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10081 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 10140

Qy        241 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10141 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 10200

Qy        301 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTTATATTGCTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10201 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTTATATTGCTG 10260

Qy        361 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10261 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 10320

Qy        421 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10321 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 10380

Qy        481 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10381 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 10440

Qy        541 AAAAAGATTCCATACTTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10441 AAAAAGATTCCATACTTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT 10500

Qy        601 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10501 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT 10560

Qy        661 GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10561 GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT 10620

Qy        721 ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10621 ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA 10680

Qy        781 TCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10681 TCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTG 10740

Qy        841 ATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10741 ATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTA 10800
```

```
Qy        901 TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10801 TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTG 10860

Qy        961 TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10861 TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT 10920

Qy       1021 GGAGTTGTTTATGGATTTGCAAAAATTAGATAA 1053
              |||||||||||||||||||||||||||||||||
Db      10921 GGAGTTGTTTATGGATTTGCAAAAATTAGATAA 10953
```

```
RESULT 4
AF285969
LOCUS       AF285969               13650 bp    DNA     linear   BCT 02-AUG-2001
DEFINITION  Salmonella enterica O antigen gene cluster, complete sequence.
ACCESSION   AF285969
VERSION     AF285969.1  GI:15076983
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 13650)
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     The E. coli O111 and S. enterica O35 gene clusters: Gene clusters
            encoding the same colitose containing O antigen are highly
            conserved
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 13650)
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2000) Department of Microbiology, The University
            of Sydney, Sydney, NSW 2006, Australia
FEATURES             Location/Qualifiers
     source          1. .13650
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /serovar="Adelaide"
                     /db_xref="taxon:28901"
                     /note="serogroup: O35"
     gene            1321. .2460
                     /gene="wbdH"
     CDS             1321. .2460
                     /gene="wbdH"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glycosyl transferase WbdH"
                     /protein_id="AAK83009.1"
                     /db_xref="GI:15076984"
                     /translation="MKKIVIIGNIASMMLRFRKELIINLVSQGDDVYCLANDFSPEDL
                     KLLSSWGGKGIKYTLNSKGINPFKDILSVIELKNILDKISPDIVFSYFVKPVIFGTIA
                     SKLSRVPRIVGMIEGLGNAFTYYKGKQSIKTRIIKWMQIFLYKLALPMLNDLVLLNKD
                     DKKDLIERYNIKVNVTILGGIGLNLNDFSYKEPLTDKVTFIFIARLLREKGVFEYIEA
                     AKRVKNIYPNVEFVILGGFENNNPFSLQQHEIELLRKEHDLIYPGHVENVQDWLEKSS
                     VFVLPTSYREGVPRVIQEAMAIGRPVITTNVPGCKDIINDGINGFLVPPFESELLVKK
                     MIYFIDNRSKILEMGLSGRLFAEKNFDAMEKNKTLASIIKANHEF"
     gene            2511. .3629
                     /gene="gmd"
     CDS             2511. .3629
                     /gene="gmd"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-D-mannose dehydratase Gmd"
                     /protein_id="AAK83010.1"
                     /db_xref="GI:15076985"
                     /translation="MNKVALITGITGQDGSYLAELLLEKGYEVHGIKRRASSFNTERV
                     DHIYQDIHLNNPKLYLHYGDLTDTSNLTRILQEVQPDEVYNLGAMSHVAVSFESPEYT
                     ADVDALGTLRLLEAIRILGLEKKTKFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
                     KLYAYWITVNYRESYGMFACNGILFNHESPRRGETFVTRKITRGIANIAQGLDKCLYL
                     GNMDSLRDWGHAKDYVKMQWMMLQQEAPEDFVIATGVQYSVRDFVTMAAKQVGIELVF
                     EGEGINEKGIVASVNGNDAKAVKAGDVIISVDPRYFRPAEVETLLGDPTNAHKKLGWS
                     PEITLCEMVKEMVSSDLAIAKKNVLLKTNNIPTNIPQE"
     gene            3698. .4147
                     /gene="gmm"
     CDS             3698. .4147
                     /gene="gmm"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-mannosyl hydrolase Gmm"
                     /protein_id="AAK83011.1"
                     /db_xref="GI:15076986"
                     /translation="MFIPSDKFREVIKLTPLVSIDILIENENGEYLFGLRNNRPAKNQ
                     FFVPGGRIRKSEFIENAFKRISFAELGKEYDISESLFNGIWEHFYEESFFSDADATHY
                     IVLCYKLKVLKNELNLPADQHCEYIWISEDKISNLNNIHKYSKDYFL"
     gene            4181. .5626
                     /gene="manC"
     CDS             4181. .5626
                     /gene="manC"
                     /codon_start=1
                     /transl_table=11
                     /product="mannose-1-phosphate guanylyltransferase ManC"
                     /protein_id="AAK83012.1"
                     /db_xref="GI:15076987"
                     /translation="MSQCLYPVVIAGGTGSRLWPLSRVLYPKQFLNLVGNSTMLQTTI
                     ARLNGIKCENPIVICNEDHRFIVAEQLRQIDKLTKNIILEPKGRNTAPAIALAAFIAQ
                     KKKPKDDPLLLVLAADHSINNEQAFREAIIKAIPYADAGKLVMFGIIPSVAHTGYGYI
                     KRSVPVDADKDSTAYYVANFVEKPNIQKAQEYIMSEGYYWNSGMFLFRASKYLGELKK
                     FRPDIYNKCESATATANIDMDFVRINETEFINCPEESIDYAVMEKTKDAVVLPIDVGW
                     SDVGSWSSLWDISQKDVHGNVCQGDVINHDSEDSFIYSESSLVATVGVSNLVIVQTKD
                     AVLVADRDKVQNVKNIVDSLKQGGRAEYYMHREVFRPWGKYDAIDQGDRYKVKKIIVK
                     PGEGLDLRMHHHRAEHWIVVSGTAKVSLGNKIKLLVPNESIYIPQGAEYSLENPGVIP
                     LHLIEVSSGDYLESDDIVRFTDRYNSKKFLK"
     gene            5639. .7015
                     /gene="manB"
     CDS             5639. .7015
                     /gene="manB"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphomannomutase ManB"
                     /protein_id="AAK83013.1"
                     /db_xref="GI:15076988"
                     /translation="MNKLTCFKAYDIRGRLGVELDDDIAYRIGRAYGELFKPQTVVVG
                     GDTRLTSESLKKSLSNGLCDAGVNVLDLGMCGTEEIYFSTWHLGVDGGIEVTASHNPI
                     DYNGMKLVTKNARPISSDTGLKDIQTLVELNDFEKINLEKTGNITKYSTRDAYIKHLM
                     GYVNLQKIKKIKIVVNSGNGAAGPVIDDIEKYFLQNNIPVTFVKINNTPDGHFPNGIP
                     NPLLPECREDTSKAVIAHSADFGIAFDGDFDRCFFFDEKGRFIEGYYIVGLLAEVFLE
                     KYPNAKIIHDPRLIWNTIDIVQNYGGKPIMTKTGHAYIKQRMREEDAVYGGEMSAHHY
                     FKDFEYCDSGMIPWILICELLSIRNKNLGELVYNCINAWPASGEINCTLSNPQSEIDK
                     LFFYYKNSAISVDYTDGLSMEFSDWRFNVRCSNTEPVVRLNVESKSNVLLMQEKTREI
                     LSFIGKCV"
     gene            7040. .7963
                     /gene="wbdJ"
     CDS             7040. .7963
                     /gene="wbdJ"
                     /codon_start=1
                     /transl_table=11
                     /product="WbdJ"
                     /protein_id="AAK83014.1"
                     /db_xref="GI:15076989"
                     /translation="MKVLLTGSSGMVGKNILEHASAKDYEILTPTSSDLNLLVIDEIE
                     KFMLANKPECIIHAAGLVGGIHANINRPFDFLANNLLMGLNLVTIAKKLGIKKFLNLG
                     SSCMYPRNFEEAIPEKALLTGELEETNEGYAIAKVAVAKACEYISRENSNYYYKTIIP
                     CNLYGKYDKFGDDSSHMIPAVIKKIHHAKANNIPEIEIWGDGNSRREFMYAEDLADLV
                     FYIIPKIESIPNMVNAGLGYDYTVNDYYKIIAEEIGYTGTFFHDLTKPTGMKRKLVDI
                     SLLNKIGWSSHFNLRDGIRKTYNYYLENQYK"
     gene            7960. .9126
                     /gene="wbdK"
     CDS             7960. .9126
                     /gene="wbdK"
                     /codon_start=1
                     /transl_table=11
                     /product="WbdK"
                     /protein_id="AAK83015.1"
```

```
                     /db_xref="GI:15076990"
                     /translation="MITYPLASNTWDENEYAAIQSVIDSKMFTMGKNVELYEKNFADL
                     FGSKYAVMVSSGSTANLLMIAALFFTHKPKLKRGDEIIVPAVSWSTTYYPLQQYGLKV
                     KFVDINKETLNIDVDSLKNAISDKTRAILAVNLLGNPNDFSTIHEIISNKDIILLEDN
                     CESMGATFQNKQAGTFGIMGSFSSFYSHHIATMEGGCVVTDDEELYHILLCLRAHGWT
                     RNLPKENMVTGTKSDDIFEESFKFVLPGYNVRPLEMSGAIGIEQLKKLPGFISTRRSN
                     AQYFVDKFKDHPFLDIQKEIGESSWFGFSLIIKQDTGIERKSLINKLITSGIECRPIV
                     TGNFLKNKRVLSYFDYSVHDTVVNAEYIDKNGFFVGNHQIPLLNEIDFLRKVLK"
gene                 9178. .10431
                     /gene="wzx"
CDS                  9178. .10431
                     /gene="wzx"
                     /codon_start=1
                     /transl_table=11
                     /product="O antigen flippase Wzx"
                     /protein_id="AAK83016.1"
                     /db_xref="GI:15076991"
                     /translation="MKKFLAFGYSKVLPPIIEQFINPICIFVITPLILNHLGKQSYGN
                     WILLITIVSFSQLICGGCSAWIAKVIAEQRATSDYAGKGALRKIAYNLSIIMLSFAIL
                     ILLILLSLCVFDVARDNTEFLYAIAICGFFQEVDNLFNGALKGFERFNVSCLFEVITR
                     VLWAFIVIYGIQENSVLYYTCLAFILKGTLKYILVCLNISGKFIYPNFDKAGIVNLLN
                     ESKWMFLQLTGGISLGLFDRLVIPLILSVNKLSSYVPCLQLVQLMFTISASANQILLP
                     MFARMKASNTFPSNCFFKIQLVSLSSALPCLVLLFFGYDILSIWISPAFASENFKLMQ
                     ILSISYILLSMMTSFHFLLLGLGKSKLVANLNLIAGLTLAASTLIAAHYGLYAISVVK
                     IIYPLFQFYYLYVAFFYFKREKNVY"
gene                 10421. .11485
                     /gene="wzy"
CDS                  10421. .11485
                     /gene="wzy"
                     /codon_start=1
                     /transl_table=11
                     /product="O antigen polymerase Wzy"
                     /protein_id="AAK83017.1"
                     /db_xref="GI:15076992"
                     /translation="MSIDFLFSITEIAVVFSCAIYIFVQCLLMHKIYLDKVVLILLCL
                     LLCLVIIQLPELNEQGLIDAVKLLLPLLIVFIAFQKPKLCLWIIIALLFFNSTLNFWH
                     FKVFDKFSSFPFTFFILLYFLYKLVNDQPPRQRNKKFYVVIFIFILIDVTQSLLISYR
                     GQILYSLICVIMLVLRINLRKTIPYIFLTLPFIYVFIMTFIGFNYFNKDVVFFEPTAS
                     NIERTGMIYYLVIHANDYIFHGMGTLHFLNEGGQYKMLYGFPSLIPNDPHDFLLRFFI
                     SIGMMGALIYHCIFFVFFKNISSLLRQKNSYFVIVSCLLLLQVVLFYTLNPFNAFNRL
                     ICGLTIGVIYGFSKSRQIYM"
gene                 11451. .12344
                     /gene="wbdL"
CDS                  11451. .12344
                     /gene="wbdL"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glycosyl transferase WbdL"
                     /protein_id="AAK83018.1"
                     /db_xref="GI:15076993"
                     /translation="MDFQRVDRYICSESFNTPLISIIIATFNSELVIANCLQSVISQS
                     YKNIEVIIMDGGSYDKTLDIARSFKDERIKIVSEKDRGIYDAWNKAVDLSSGDWIAFI
                     GSDDVYYHTDAITSLVKGAAISNGAPVVYGRTAHEGPNKEISGFSGSEWYNLKGFKFN
                     YYKCNLPLPIMSAIYSRDFFKDQRFDIKLKIVADADWFLRCFIKWRKENSPYFIDDIT
                     PVVRMGYGGVSTDISSQIKTTRESFIVRKKNNISCFNMQLILRYIKIMVMVSIKNIFG
                     NNIYNSMHNGYHFLKKIKNKI"
gene                 12341. .13465
                     /gene="wbdM"
CDS                  12341. .13465
                     /gene="wbdM"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glycosyl transferase WbdM"
                     /protein_id="AAK83019.1"
                     /db_xref="GI:15076994"
                     /translation="MKILYVITGLTCGGAEHLMMQLADQMLLRGHDVNIICLSGISEI
                     QPTQKHNIHYIRMEKNVRGFVKALFQVRKIITVIKPDIIHSHMFHANIFSRIIRILTP
                     SVPLICTAHNKNEGGYARMLCYRLSDFLASITTNVSQEAVQEFIAKKAAPKNKIIEVP
                     NFVNTTNFSFDLKARKERRNFFGIKDNTLLLLAVGRLVDAKDYPNLLNAINYLTLSES
                     TGFDNFILFIAGDGVLKSKLVDLVNKLNLADKVFFLGRRSDIKELMCAADLFVLSSQW

Query Match          65.1%;  Score 685.6;  DB 1;  Length 13650;
Best Local Similarity 78.2%;  Pred. No. 6.4e-81;

  Matches  823;  Conservative    0;  Mismatches  229;  Indels    0;  Gaps    0;

Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              |||||||||||||| ||||||||||||||||| || ||| |||||||||||| |||||
Db      10421 ATGTCTATTGATTTTCTTTTTTCAATTACTGAGATTGCAGTTGTTTTTTCTTGTGCTATT 10480

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
              || |||||  ||||| |||||||||  | ||| ||||||||||||  | ||||||||
Db      10481 TATATATTTGTTCAATGCTTGTTAATGCACAAGATTTATTTAGATAAAGTTGTTTTAATT 10540

Qy        121 CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 180
              || || || || ||| |||  || || || |||||||||||| || |||||||    | |||
Db      10541 CTATTGTGTTTACTCCTTTGCTTGGTCATTATTCAACTTCCCGAACTTAATGAGCAAGGT 10600

Qy        181 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 240
              ||| | ||| || ||||| ||| | | |||||| | || || |||||||||||||||||
Db      10601 TTGATAGATGCTGTAAAGCTATTATTACCTTTACTAATAGTATTTATCGCTTTTCAAAAA 10660

Qy        241 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 300
              |||||||||||||||||| | |||||||||| ||||| || || ||  |  | |||||
Db      10661 CCGAAATTATGCTTGTGGATCATTATTGCATTATTGTTCTTCAATTCCACGCTAAATTTC 10720

Qy        301 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTG 360
              |    | || ||    || ||||| ||||||||||||||||||||| ||||||||||  |
Db      10721 TGGCACTTCAAAGTTTTTGATAAATTTAGCTCATTTCCTTTTACATTTTTTATATTATTA 10780

Qy        361 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 420
              | ||  |||| | ||||| | || | |  ||     | | ||||||||||||||||
Db      10781 TATTTTTTGTACAAATTGGTCAATGATCAACCTCCAAGGCAGAGGAATAAAAAATTTTAC 10840

Qy        421 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 480
              |    | ||||| | |||||||||| ||  |||  || ||||| ||||||| ||| ||
Db      10841 GTTGTAATTTTCATATTTATATTAATTGATGTAACACAATCATTATTAATAAGTTACAGA 10900

Qy        481 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 540
              || |||||| |||| ||  |||| ||   |  | || ||||| || |   ||||||||||
Db      10901 GGACAGATTCTATACTCTTTAATATGTGTTATAATGCTTGTATTGAGGATTAATTTAAGG 10960

Qy        541 AAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT 600
              ||||  || |||||  |||||  |||   | ||  || |||||||  |||| ||| ||| |
Db      10961 AAAACAATCCCATATATTTTTCTAACTTTACCGTTTATATATGTTTTTATAATGACTTTT 11020

Qy        601 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT 660
              |||||||||||||| || |||||||  |||  |||||||||||||| ||||||||||||
Db      11021 ATTGGTTTTAATTACTTTAATAAAGACGTGGTTTTTTTTGAACCTACGGCAAGTAATATT 11080

Qy        661 GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT 720
              |||||||||||||||||||||||||||||||||  ||      |||||||||||| ||| |||
Db      11081 GAACGTACGGGGATGATATATTATTTGGTTATCCATGCAAATGATTATATATTTCACGGT 11140

Qy        721 ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA 780
              |||||||| ||| |||||||||| | || || |||||||| | |||||||||| ||||
Db      11141 ATGGGGACGTTACATTTCTTAAATGAGGGGGGGCAATATAAAATGTTATATGGATTTCCT 11200

Qy        781 TCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTG 840
              ||  |||| || ||||| |||||||||||||||||| | |||||||||||||||||| ||
Db      11201 TCTCTAATACCAAATGATCCTCATGATTTTTTATTAAGATTCTTTATAAGTATTGGTATG 11260

Qy        841 ATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTA 900
              || || || ||| | ||||||| |||||||||||||||||||  | |||||||  |||
Db      11261 ATGGGGGCGTTGATCTATCATTGTATATTTTTTGTTTTTTTTTAAAAATATATCTTCGTTA 11320

Qy        901 TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTG 960
              ||    | | |||    |  ||| |  | |||||||||| ||| | ||||||||||||
Db      11321 TTGCGTCAAAAAAACAGTTATTTTGTCATAGTAAGTTGTTTATTATTATTACAAGTTGTG 11380

Qy        961 TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT 1020
              ||  |||||||| |||||||||| |||| |||||||||| ||||||||||||||  |
Db      11381 TTGTTTTATACACTAAACCCTTTTAATGCATTTAATCGATTAATTTGCGGGCTTACAATC 11440

Qy       1021 GGAGTTGTTTATGGATTTGCAAAAATTAGATA 1052
              |||||  |||||||||| |||| | |||| |
Db      11441 GGAGTCATTTATGGATTTTCAAAGAGTAGACA 11472
```

```
RESULT 5
AJ840631
LOCUS       AJ840631               1328 bp    DNA     linear   PLN 22-SEP-2004
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, right border, clone
            603D05.
ACCESSION   AJ840631
VERSION     AJ840631.1  GI:52544837
KEYWORDS    right border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
   PUBMED   12446565
REFERENCE   2  (bases 1 to 1328)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2004) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .1328
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="603D05"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Wassilewskija"
     misc_feature    1. .1328
                     /note="T-DNA flanking sequence
                     right border"
ORIGIN

  Query Match             12.7%;  Score 133.8;  DB 15;  Length 1328;
  Best Local Similarity   42.8%;  Pred. No. 1.6e-08;
  Matches  444;  Conservative    0;  Mismatches  593;  Indels    0;  Gaps    0;

Qy          2 TGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTT 61
              | | | |   |||  ||||||   |         |      || |||||| ||    | |||
Db        209 TTTTTTTNTTTTTTTTTTTTTTTTTNNTNTTTTTTTTTNTTTTTTTTTTTTTTTTTTTT 268

Qy         62 ACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTC 121
                |  ||  |   | ||| ||  |         | | ||     |     | ||||  ||
Db        269 TTTTNNTTTTTTTNTTTTTTTTTTTTTTTTTTTNTTTTTTNTTTTNTNTTTNTTTTTTTT 328

Qy        122 TTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTT 181
              |||| |  ||  | |||||||||  |  |  ||    ||  |    ||  | |      ||
Db        329 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTT 388

Qy        182 TGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAAC 241
              |  |   || |||    || |   |   ||| ||  |  | ||| |   ||||
Db        389 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 448

Qy        242 CGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTT 301
```

```
                  || |  || |   || || ||   || || |||||    | |   |||  |||||
Db        449 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTT 508

Qy        302 TATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGT 361
              | | |||         ||   |   |||   |  |||  ||||  |||| || | ||  | |
Db        509 TTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTT 568

Qy        362 TTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACG 421
              |||  || |||   ||     |   |  |||     | | |     |     ||||
Db        569 TTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTNTTTTTTTTTTTTTNTTTTTTTT 628

Qy        422 CGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGG 481
                ||  ||||| | ||| | ||  |     |  |     |  || ||  |   || |
Db        629 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTN 688

Qy        482 GGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAA 541
                  |||| | ||    |  |||   |  |   |  || | |||    ||  |||
Db        689 TTTTTTTTTNTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 748

Qy        542 AAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATA 601
                  |    | ||||||||  |  |     ||||  | | |  || || |   || |
Db        749 TTTTTTNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTT 808

Qy        602 TTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTG 661
              ||  ||||  || |||    |       |  |||||||     |       |  | ||
Db        809 TTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTNTTTTTNTTTTTNTTTTT 868

Qy        662 AACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTA 721
                |     | | | || ||| |||       ||  |  || | | ||    |  |
Db        869 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 928

Qy        722 TGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCAT 781
              |       ||   ||| ||   |           | |     | | |    ||   |
Db        929 TTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTNTTTTTTTTTTTTT 988

Qy        782 CATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGA 841
                ||  ||  |  |      |      |||||| ||    || ||| |   | ||  |
Db        989 TTTTTTTTTTNTTTTTTTTTNNTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTNNT 1048

Qy        842 TAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTAT 901
              |       |  ||| |  || | | ||| || ||||||||      | | |   || |
Db       1049 TTTTTTNTNTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTNTTTNNTTTTTTNNTTTTT 1108

Qy        902 TATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGT 961
              | | |       |  | |||  || || |   || ||| ||    ||    || |
Db       1109 TTTTTTTTTTTNNTTTTTTTTTTTTTTTTTTTNTTTNTTTNTTTTNNTTTTNNTTTNTNN 1168

Qy        962 TAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTTG 1021
              |  |||      ||      ||||  |   |||  |   ||  |||      |    ||
Db       1169 TNNTTTNNTTTTTTTTTNTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTNTTTNTNNNNTTT 1228

Qy       1022 GAGTTGTTTATGGATTT 1038
                || ||  |   |||
Db       1229 TTTTTTTTNNTTTTTTT 1245
```

```
RESULT 6
AC152242
LOCUS       AC152242               282289 bp    DNA     linear   HTG 01-JUL-2005
DEFINITION  Bos taurus clone CH240-7O10, *** SEQUENCING IN PROGRESS ***, 38
            unordered pieces.
ACCESSION   AC152242
VERSION     AC152242.3  GI:68227126
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 282289)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
```

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 282289)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 282289)
AUTHORS Cow Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jun 26, 2005 this sequence version replaced gi:58038019.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: FAGP
Center clone name: CH240-7O10
---------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 265527 bases at least Q40
Consensus quality: 268628 bases at least Q30
Consensus quality: 271489 bases at least Q20
Estimated insert size: 269724; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 18156: contig of 18156 bp in length
* 18157 18206: gap of 50 bp
* 18207 50971: contig of 32765 bp in length
* 50972 51021: gap of 50 bp
* 51022 62914: contig of 11893 bp in length
* 62915 62964: gap of 50 bp
* 62965 65254: contig of 2290 bp in length
* 65255 65304: gap of 50 bp
* 65305 66643: contig of 1339 bp in length
* 66644 66693: gap of 50 bp
* 66694 72399: contig of 5706 bp in length
* 72400 72821: gap of 422 bp
* 72822 74487: contig of 1666 bp in length
* 74488 74537: gap of 50 bp
* 74538 80785: contig of 6248 bp in length
* 80786 80835: gap of 50 bp
* 80836 87779: contig of 6944 bp in length
* 87780 87879: gap of unknown length
* 87880 92358: contig of 4479 bp in length
* 92359 92408: gap of 50 bp
* 92409 97196: contig of 4788 bp in length
* 97197 97246: gap of 50 bp
* 97247 104787: contig of 7541 bp in length
* 104788 104837: gap of 50 bp
* 104838 106100: contig of 1263 bp in length
* 106101 106200: gap of unknown length
* 106201 113418: contig of 7218 bp in length
* 113419 113606: gap of 188 bp
* 113607 115164: contig of 1558 bp in length
* 115165 115214: gap of 50 bp
* 115215 126597: contig of 11383 bp in length
* 126598 126697: gap of unknown length
* 126698 130816: contig of 4119 bp in length
* 130817 130866: gap of 50 bp
* 130867 133210: contig of 2344 bp in length
* 133211 134248: gap of 1038 bp
* 134249 137021: contig of 2773 bp in length
* 137022 138232: gap of 1211 bp
* 138233 139433: contig of 1201 bp in length
* 139434 139533: gap of unknown length
* 139534 144869: contig of 5336 bp in length
* 144870 144919: gap of 50 bp
* 144920 151314: contig of 6395 bp in length
* 151315 151364: gap of 50 bp
* 151365 174159: contig of 22795 bp in length
* 174160 174209: gap of 50 bp

```
          *   174210   216578: contig of 42369 bp in length
          *   216579   216628: gap of 50 bp
          *   216629   223479: contig of 6851 bp in length
          *   223480   223529: gap of 50 bp
          *   223530   229963: contig of 6434 bp in length
          *   229964   230118: gap of 155 bp
          *   230119   233635: contig of 3517 bp in length
          *   233636   233860: gap of 225 bp
          *   233861   245820: contig of 11960 bp in length
          *   245821   245870: gap of 50 bp
          *   245871   253648: contig of 7778 bp in length
          *   253649   253748: gap of unknown length
          *   253749   254859: contig of 1111 bp in length
          *   254860   254959: gap of unknown length
          *   254960   256040: contig of 1081 bp in length
          *   256041   256140: gap of unknown length
          *   256141   257392: contig of 1252 bp in length
          *   257393   257492: gap of unknown length
          *   257493   258889: contig of 1397 bp in length
          *   258890   258989: gap of unknown length
          *   258990   260305: contig of 1316 bp in length
          *   260306   260405: gap of unknown length
          *   260406   261931: contig of 1526 bp in length
          *   261932   262031: gap of unknown length
          *   262032   264065: contig of 2034 bp in length
          *   264066   264165: gap of unknown length
          *   264166   271493: contig of 7328 bp in length
          *   271494   271593: gap of unknown length
          *   271594   282289: contig of 10696 bp in length.
FEATURES             Location/Qualifiers
     source          1. .282289
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /clone="CH240-7O10"
     gap             18157. .18206
                     /estimated_length=50
     gap             50972. .51021
                     /estimated_length=50

  Query Match             11.5%;  Score 121.4;  DB 14;  Length 282289;
  Best Local Similarity   39.1%;  Pred. No. 6.2e-08;
  Matches  401;  Conservative    0;  Mismatches  623;  Indels    2;  Gaps    1;

Qy         19 TTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTACATATTTACTCAATGT 78
Db      50036 TTTGGTTTTTCTACACGAGACTTTTTTTTTTTTTTTTTTTTCNNCCNNTTTTTTTTTTT 50095

Qy         79 TTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTTATGCTTGCTCTTT 138
Db      50096 TNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNNNTTTTNNTTTTTTTTTTNNTTTTTTT 50155

Qy        139 TTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGTCGATTCTTTAAAG 198
Db      50156 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNNTTTT 50215

Qy        199 TTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAAATTATGCTTGTGG 258
Db      50216 TTTTTTTTTTTTTTTTTTTTTTNNTTTTNNTTTNNTTNNTTTTTTNNTTNNTTTTTTT 50275

Qy        259 GTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTAAAGACATTC 318
Db      50276 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNN 50335

Qy        319 GATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACTTGTTTAGATTG 378
Db      50336 TNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTT 50395

Qy        379 GGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGATTTTTCTCTTT 438
Db      50396 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTNNNTTT 50455

Qy        439 ATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCAGATTTTATATTCC 498
Db      50456 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTTTTTTTTT 50515

Qy        499 GTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGATTCCATACTTT 558
Db      50516 TTTTTTTTTNNTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTTTNNTTTTT 50575

Qy        559 TTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGGTTTTAATTATTTC 618
Db      50576 TTTTTNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTNNTTTTTTNNNTTTTNNTTTT 50635

Qy        619 AATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTACGGGGATGATA 678
Db      50636 TTTTNNNNNNNTTTTTTNNTTTTTTTTTTTTTTTTTNNNTTTNNNTTTTTTTTTTNNTT 50695

Qy        679 TATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATGGGGACATTAAATTTC 738
Db      50696 TTTTTTTT--TTTTNNTTTTNNNNTTTTTTTTTCCCTTNNNTTTTTNNTTTTTTTTT 50753

Qy        739 TTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATTAATTCCTAATGAC 798
Db      50754 TNNTTTTTNNNNTTTTTTTTTTTTTTTTTTTTNNNNTTNNTTTTTTTTTTTTTTTTTT 50813

Qy        799 CCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATAGGAGCATTGGTTTAT 858
Db      50814 TTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTTTTTTTTNNTTTNNTTTNNTTTTTTT 50873

Qy        859 CATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTATTATATGAGAGAAATGCT 918
Db      50874 TTTTTTNNTTTTTTTTTNNTTTTTTTTTNNNTTTTTTTTTTTTTTTTTTTNNNTTTTTNN 50933

Qy        919 CCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAATTTATACATTAAAC 978
Db      50934 NTTTTTTTTTTTTTNNTTTTTTTTTTNNNTTTCCAGGNNNNNNNNNNNNNNNNNNNN 50993

Qy        979 CCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTTGGAGTTGTTTATGGATTT 1038
Db      50994 NNNNNNNNNNNNNNNNNNNNNNNNNNNAGGGGCGGGCGGGGGGGCATACAAAAAGAGGA 51053

Qy       1039 GCAAAA 1044
Db      51054 GCCAAA 51059
```

```
RESULT 7
AC138074
LOCUS       AC138074               66993 bp    DNA     linear   HTG 12-DEC-2002
DEFINITION  Homo sapiens chromosome 18 clone RP13-907D8 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC138074
VERSION     AC138074.1  GI:26553381
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 66993)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP13-907D8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 66993)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
```

```
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L28874
                Center clone name: 907_D_8
            --------------.
            * NOTE: This record contains 55 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1     1116: contig of 1116 bp in length
            *     1117     1216: gap of 100 bp
            *     1217     2336: contig of 1120 bp in length
            *     2337     2436: gap of 100 bp
            *     2437     3561: contig of 1125 bp in length
            *     3562     3661: gap of 100 bp
            *     3662     4781: contig of 1120 bp in length
            *     4782     4881: gap of 100 bp
            *     4882     6001: contig of 1120 bp in length
            *     6002     6101: gap of 100 bp
            *     6102     7244: contig of 1143 bp in length
            *     7245     7344: gap of 100 bp
            *     7345     8458: contig of 1114 bp in length
            *     8459     8558: gap of 100 bp
            *     8559     9640: contig of 1082 bp in length
            *     9641     9740: gap of 100 bp
            *     9741    10846: contig of 1106 bp in length
            *    10847    10946: gap of 100 bp
            *    10947    12103: contig of 1157 bp in length
            *    12104    12203: gap of 100 bp
            *    12204    13320: contig of 1117 bp in length
            *    13321    13420: gap of 100 bp
            *    13421    14520: contig of 1100 bp in length
            *    14521    14620: gap of 100 bp
            *    14621    15773: contig of 1153 bp in length
            *    15774    15873: gap of 100 bp
            *    15874    16991: contig of 1118 bp in length
            *    16992    17091: gap of 100 bp
            *    17092    18240: contig of 1149 bp in length
            *    18241    18340: gap of 100 bp
            *    18341    19476: contig of 1136 bp in length
            *    19477    19576: gap of 100 bp
            *    19577    20721: contig of 1145 bp in length
            *    20722    20821: gap of 100 bp
            *    20822    21942: contig of 1121 bp in length
            *    21943    22042: gap of 100 bp
            *    22043    23141: contig of 1099 bp in length
            *    23142    23241: gap of 100 bp
            *    23242    24372: contig of 1131 bp in length
            *    24373    24472: gap of 100 bp
            *    24473    25627: contig of 1155 bp in length
            *    25628    25727: gap of 100 bp
            *    25728    26888: contig of 1161 bp in length
            *    26889    26988: gap of 100 bp
            *    26989    28030: contig of 1042 bp in length
            *    28031    28130: gap of 100 bp
            *    28131    29303: contig of 1173 bp in length
            *    29304    29403: gap of 100 bp
            *    29404    30554: contig of 1151 bp in length
            *    30555    30654: gap of 100 bp
            *    30655    31751: contig of 1097 bp in length
            *    31752    31851: gap of 100 bp
            *    31852    32966: contig of 1115 bp in length
            *    32967    33066: gap of 100 bp
            *    33067    34113: contig of 1047 bp in length
            *    34114    34213: gap of 100 bp
            *    34214    35329: contig of 1116 bp in length
            *    35330    35429: gap of 100 bp
            *    35430    36574: contig of 1145 bp in length
            *    36575    36674: gap of 100 bp
            *    36675    37828: contig of 1154 bp in length
            *    37829    37928: gap of 100 bp
            *    37929    39063: contig of 1135 bp in length
            *    39064    39163: gap of 100 bp
            *    39164    40289: contig of 1126 bp in length
            *    40290    40389: gap of 100 bp
            *    40390    41471: contig of 1082 bp in length
            *    41472    41571: gap of 100 bp
            *    41572    42634: contig of 1063 bp in length
            *    42635    42734: gap of 100 bp
            *    42735    43904: contig of 1170 bp in length
            *    43905    44004: gap of 100 bp
            *    44005    45128: contig of 1124 bp in length
            *    45129    45228: gap of 100 bp
            *    45229    46358: contig of 1130 bp in length
            *    46359    46458: gap of 100 bp
            *    46459    47595: contig of 1137 bp in length
            *    47596    47695: gap of 100 bp
            *    47696    48785: contig of 1090 bp in length
            *    48786    48885: gap of 100 bp
            *    48886    49960: contig of 1075 bp in length
            *    49961    50060: gap of 100 bp
            *    50061    51132: contig of 1072 bp in length
            *    51133    51232: gap of 100 bp
            *    51233    52356: contig of 1124 bp in length
            *    52357    52456: gap of 100 bp
            *    52457    53556: contig of 1100 bp in length
            *    53557    53656: gap of 100 bp
            *    53657    54782: contig of 1126 bp in length
            *    54783    54882: gap of 100 bp
            *    54883    56008: contig of 1126 bp in length
            *    56009    56108: gap of 100 bp
            *    56109    57307: contig of 1199 bp in length
            *    57308    57407: gap of 100 bp
            *    57408    58529: contig of 1122 bp in length
            *    58530    58629: gap of 100 bp
            *    58630    59799: contig of 1170 bp in length
            *    59800    59899: gap of 100 bp
            *    59900    60983: contig of 1084 bp in length
            *    60984    61083: gap of 100 bp
            *    61084    62193: contig of 1110 bp in length
            *    62194    62293: gap of 100 bp
            *    62294    63397: contig of 1104 bp in length
            *    63398    63497: gap of 100 bp
            *    63498    64574: contig of 1077 bp in length
            *    64575    64674: gap of 100 bp
            *    64675    65800: contig of 1126 bp in length
            *    65801    65900: gap of 100 bp
            *    65901    66993: contig of 1093 bp in length.
FEATURES                Location/Qualifiers
     source             1. .66993
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP13-907D8"
                     /clone_lib="RPCI-13 Human Female BAC"
     gap             1117. .1216
                     /estimated_length=100
     gap             2337. .2436
                     /estimated_length=100
     gap             3562. .3661
                     /estimated_length=100
     gap             4782. .4881
                     /estimated_length=100
     gap             6002. .6101
                     /estimated_length=100
     gap             7245. .7344
                     /estimated_length=100
     gap             8459. .8558
                     /estimated_length=100
     gap             9641. .9740
                     /estimated_length=100
     gap             10847. .10946
                     /estimated_length=100
     gap             12104. .12203
                     /estimated_length=100
     gap             13321. .13420
                     /estimated_length=100

  Query Match             11.3%;  Score 119.4;  DB 14;  Length 66993;
  Best Local Similarity   40.0%;  Pred. No. 2.1e-07;
  Matches  399;  Conservative    0;  Mismatches  593;  Indels    5;  Gaps    1;

Qy         10 GATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTACATATTT 69
Db      56299 GAATCATGTTTCTCACNCTTNATNNTNNNNNNNTNTTATNNTTNNNTTNTTTTTTTTTTN 56358

Qy         70 ACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTTATGC 129
Db      56359 NTTTNTTTCTTNNTTTTNTTTTTTTTNTTTTTTTTTTTTTNNNNTNTTTNNTNNTTTTNNN 56418

Qy        130 TTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGTCGAT 189
Db      56419 TTTTNNTTTTTNNTTTTNTTTTTTTTTTTTTTTNNCCCTTTTTT-----TTTTNTTTTTT 56473

Qy        190 TCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAAATTA 249
Db      56474 TTTNTNTTTTTTTNTTTTTTTTTTTTTNTTNTTTTTTTTTTTTTTTTTNNNNTNNNNTNNTT 56533

Qy        250 TGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTA 309
Db      56534 TTTTTTTTTTTTTTTNNNNTNNTTTNTNTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTT 56593

Qy        310 AAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACTTG 369
Db      56594 TTTNNTTTTTTTTTTTTTTTNTTTNTTTTTTTTTTNTTTTTTTTNTTTTTTTTTTTTTTT 56653

Qy        370 TTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGATT 429
Db      56654 TTTTNNTTTNTNTTNTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTNTTT 56713

Qy        430 TTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCAGATT 489
Db      56714 TTTTNTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTNTNTTTTTTTTTTTTTTTTTTTTTT 56773

Qy        490 TTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGATT 549
Db      56774 TTTTTTTTTTTTNTTTTTNTTTTTTTNNTTNTTTTTTTTTTTTTTNTTNTNTTTTTTTTTTT 56833

Qy        550 CCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGGTTTT 609
Db      56834 TTTTNNTNNTTNNTTTNTTTTTNTTNTTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTTT 56893

Qy        610 AATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTACG 669
Db      56894 TTTTTTNTTNNTNNTNTTTTTTTNTTTTTTNTTTTNTTNTTTNTTTTTTTTTTTTTTTNN 56953

Qy        670 GGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATGGGGACA 729
Db      56954 TTTTNTTTTTTTTTNTTTTTTTTTTNNNTTTTTTTTNTTTTTTTTTTTTTTTNNNTTTTTTT 57013

Qy        730 TTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATTAATT 789
Db      57014 TTTTTTTTTTTTTTTTTTNNNTTTTTTTNTTTTTTTTTTTTTTTTTNTNNTNNNNNTTNTT 57073

Qy        790 CCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATAGGAGCA 849
Db      57074 TNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTNTNTTTTTTT 57133

Qy        850 TTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTATTATATGAG 909
Db      57134 TTTTTTTTTTNTTTTTTTTTTTTTTTNTTTTTTTNTTTTTTTTTTTTTTNTTTTTTTTTTTT 57193

Qy        910 AGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAATTTAT 969
Db      57194 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTT 57253

Qy        970 ACATTAAACCCTTTTGATGCTTTTAATCGATTGATTT 1006
Db      57254 TTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTT 57290

RESULT 8
AJ840525
LOCUS       AJ840525               1142 bp    DNA     linear   PLN 22-SEP-2004
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, right border, clone
            598A12.
ACCESSION   AJ840525
VERSION     AJ840525.1  GI:52544731
KEYWORDS    right border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
   PUBMED   12446565
REFERENCE   2  (bases 1 to 1142)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2004) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .1142
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="598A12"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Wassilewskija"
```

```
     misc_feature    1. .1142
                     /note="T-DNA flanking sequence
                     right border"
ORIGIN

  Query Match           11.2%;  Score 118;  DB 15;  Length 1142;
  Best Local Similarity  41.8%;  Pred. No. 2e-06;
  Matches  415;  Conservative    0;  Mismatches  575;  Indels    2;  Gaps    1;

Qy         17 TTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTACATATTTACTCAAT  76
Db        148 TTNNNTNTNNNNNNNNNNNNNNNNNNTTNNTTNTTTNNTTTTTNTTTTTTTNTNTTTT 207

Qy         77 GTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTTATGCTTGCTCT 136
Db        208 TTNTTTTTTTNNTTNTTNTTTTTTTTNNTTNTTTTTTTTTTTTTTTTNTTNNNNNTTTTNT 267

Qy        137 TTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGTCGATTCTTTAA 196
Db        268 TTNTTNNNTTTTNNNTTTTNTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTN 327

Qy        197 AGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAAATTATGCTTGT 256
Db        328 NTTTTTTNTTTTTTTTTTNNNTTNNTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTNTTTT 387

Qy        257 GGGTT--ATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTAAAGAC 314
Db        388 TTTTTTNNTTTTTTTNTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTNTTTTTTTTTTN 447

Qy        315 ATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACTTGTTTAG 374
Db        448 NTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 507

Qy        375 ATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGATTTTTCT 434
Db        508 TTTTTTTTTTTTTTTTTTTTTTNTTTNTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTT 567

Qy        435 CTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCAGATTTTATA 494
Db        568 TTTTTTTTTTTTTTTTTTNTTTTNTNNTNTTTTTTTTTNTTTTTTTTTTTTTTTTTTTT 627

Qy        495 TTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGATTCCATA 554
Db        628 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 687

Qy        555 CTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGGTTTTAATTA 614
Db        688 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 747

Qy        615 TTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTACGGGGAT 674
Db        748 TTTTTTTTTTNTTTTNTTTTTTTTTTNTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTT 807

Qy        675 GATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATGGGGACATTAAA 734
Db        808 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 867

Qy        735 TTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATTAATTCCTAA 794
Db        868 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTNTTTTT 927

Qy        795 TGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATAGGAGCATTGGT 854
Db        928 TTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNT 987

Qy        855 TTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTATTATATGAGAGAAA 914
Db        988 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTT 1047

Qy        915 TGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAATTTATACATT 974
Db       1048 TTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1107

Qy        975 AAACCCTTTTGATGCTTTTAATCGATTGATTT 1006
Db       1108 TTTTTTTTTTTTTNTTTTTTTTTTTTTTTNTTT 1139

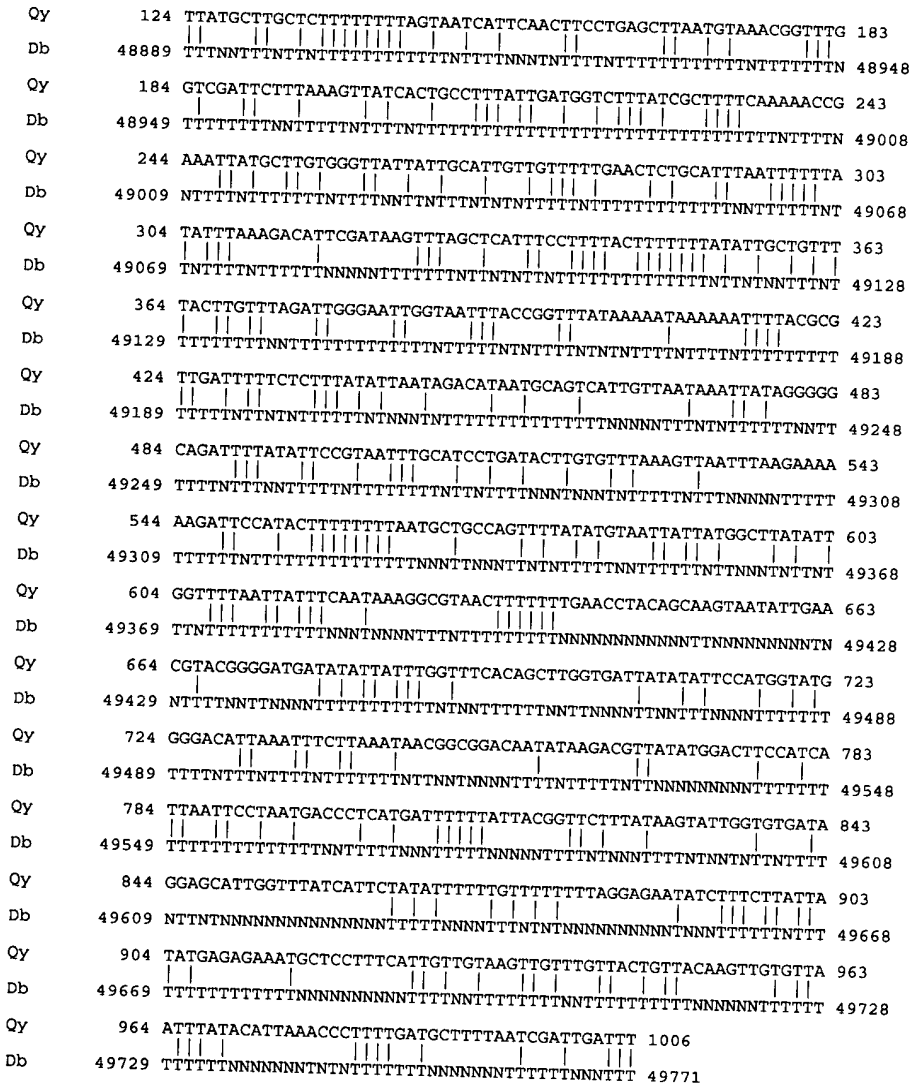

RESULT 9
AC138073
LOCUS       AC138073               93791 bp    DNA     linear   HTG 12-DEC-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-430B7 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC138073
VERSION     AC138073.1  GI:26553380
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 93791)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-430B7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 93791)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L28756
                Center clone name: 430_B_7
            --------------.
            * NOTE: This record contains 68 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1     1288: contig of 1288 bp in length
            *    1289     1388: gap of 100 bp
            *    1389     2656: contig of 1268 bp in length
            *    2657     2756: gap of 100 bp
            *    2757     4013: contig of 1257 bp in length
            *    4014     4113: gap of 100 bp
```

```
*     4114     5362: contig of 1249 bp in length
*     5363     5462: gap of 100 bp
*     5463     6748: contig of 1286 bp in length
*     6749     6848: gap of 100 bp
*     6849     8101: contig of 1253 bp in length
*     8102     8201: gap of 100 bp
*     8202     9503: contig of 1302 bp in length
*     9504     9603: gap of 100 bp
*     9604    10826: contig of 1223 bp in length
*    10827    10926: gap of 100 bp
*    10927    12171: contig of 1245 bp in length
*    12172    12271: gap of 100 bp
*    12272    13557: contig of 1286 bp in length
*    13558    13657: gap of 100 bp
*    13658    14957: contig of 1300 bp in length
*    14958    15057: gap of 100 bp
*    15058    16339: contig of 1282 bp in length
*    16340    16439: gap of 100 bp
*    16440    17747: contig of 1308 bp in length
*    17748    17847: gap of 100 bp
*    17848    19125: contig of 1278 bp in length
*    19126    19225: gap of 100 bp
*    19226    20543: contig of 1318 bp in length
*    20544    20643: gap of 100 bp
*    20644    21933: contig of 1290 bp in length
*    21934    22033: gap of 100 bp
*    22034    23288: contig of 1255 bp in length
*    23289    23388: gap of 100 bp
*    23389    24637: contig of 1249 bp in length
*    24638    24737: gap of 100 bp
*    24738    26080: contig of 1343 bp in length
*    26081    26180: gap of 100 bp
*    26181    27463: contig of 1283 bp in length
*    27464    27563: gap of 100 bp
*    27564    28816: contig of 1253 bp in length
*    28817    28916: gap of 100 bp
*    28917    30228: contig of 1312 bp in length
*    30229    30328: gap of 100 bp
*    30329    31624: contig of 1296 bp in length
*    31625    31724: gap of 100 bp
*    31725    33059: contig of 1335 bp in length
*    33060    33159: gap of 100 bp
*    33160    34450: contig of 1291 bp in length
*    34451    34550: gap of 100 bp
*    34551    35813: contig of 1263 bp in length
*    35814    35913: gap of 100 bp
*    35914    37183: contig of 1270 bp in length
*    37184    37283: gap of 100 bp
*    37284    38505: contig of 1222 bp in length
*    38506    38605: gap of 100 bp
*    38606    39921: contig of 1316 bp in length
*    39922    40021: gap of 100 bp
*    40022    41265: contig of 1244 bp in length
*    41266    41365: gap of 100 bp
*    41366    42680: contig of 1315 bp in length
*    42681    42780: gap of 100 bp
*    42781    44112: contig of 1332 bp in length
*    44113    44212: gap of 100 bp
*    44213    45557: contig of 1345 bp in length
*    45558    45657: gap of 100 bp
*    45658    47001: contig of 1344 bp in length
*    47002    47101: gap of 100 bp
*    47102    48400: contig of 1299 bp in length
*    48401    48500: gap of 100 bp
*    48501    49780: contig of 1280 bp in length
*    49781    49880: gap of 100 bp
*    49881    51183: contig of 1303 bp in length
*    51184    51283: gap of 100 bp
*    51284    52519: contig of 1236 bp in length
*    52520    52619: gap of 100 bp
*    52620    53880: contig of 1261 bp in length
*    53881    53980: gap of 100 bp
*    53981    55274: contig of 1294 bp in length
*    55275    55374: gap of 100 bp
*    55375    56691: contig of 1317 bp in length
*    56692    56791: gap of 100 bp
*    56792    58065: contig of 1274 bp in length
*    58066    58165: gap of 100 bp
*    58166    59539: contig of 1374 bp in length
*    59540    59639: gap of 100 bp
*    59640    60905: contig of 1266 bp in length
*    60906    61005: gap of 100 bp
*    61006    62302: contig of 1297 bp in length
*    62303    62402: gap of 100 bp
*    62403    63596: contig of 1194 bp in length
*    63597    63696: gap of 100 bp
*    63697    64984: contig of 1288 bp in length
*    64985    65084: gap of 100 bp
*    65085    66268: contig of 1184 bp in length
*    66269    66368: gap of 100 bp
*    66369    67636: contig of 1268 bp in length
*    67637    67736: gap of 100 bp
*    67737    69026: contig of 1290 bp in length
*    69027    69126: gap of 100 bp
*    69127    70399: contig of 1273 bp in length
*    70400    70499: gap of 100 bp
*    70500    71763: contig of 1264 bp in length
*    71764    71863: gap of 100 bp
*    71864    73116: contig of 1253 bp in length
*    73117    73216: gap of 100 bp
*    73217    74715: contig of 1499 bp in length
*    74716    74815: gap of 100 bp
*    74816    76106: contig of 1291 bp in length
*    76107    76206: gap of 100 bp
*    76207    77422: contig of 1216 bp in length
*    77423    77522: gap of 100 bp
*    77523    78820: contig of 1298 bp in length
*    78821    78920: gap of 100 bp
*    78921    80230: contig of 1310 bp in length
*    80231    80330: gap of 100 bp
*    80331    81610: contig of 1280 bp in length
*    81611    81710: gap of 100 bp
*    81711    83067: contig of 1357 bp in length
*    83068    83167: gap of 100 bp
*    83168    84434: contig of 1267 bp in length
*    84435    84534: gap of 100 bp
*    84535    85791: contig of 1257 bp in length
*    85792    85891: gap of 100 bp
*    85892    86978: contig of 1087 bp in length
*    86979    87078: gap of 100 bp
*    87079    88352: contig of 1274 bp in length
*    88353    88452: gap of 100 bp
*    88453    89751: contig of 1299 bp in length
*    89752    89851: gap of 100 bp
*    89852    91085: contig of 1234 bp in length
*    91086    91185: gap of 100 bp
*    91186    92443: contig of 1258 bp in length
*    92444    92543: gap of 100 bp
*    92544    93791: contig of 1248 bp in length.
FEATURES             Location/Qualifiers
     source          1. .93791
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"

  Query Match           11.1%;  Score 116.8;  DB 14;  Length 93791;
  Best Local Similarity   33.6%;  Pred. No. 4e-07;
  Matches  337;  Conservative    0;  Mismatches  666;  Indels    0;  Gaps    0;

Qy          4 TCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTAC 63
              | |     |||  | ||||   ||  |     |     || |||||  ||    | |||
Db      48769 TTTTNNNTTTTTTTTNTTTTTTTTTTTTTTTTTTTTNNNNTTTTTTTTTNNTTTTTTTTTTTTN 48828

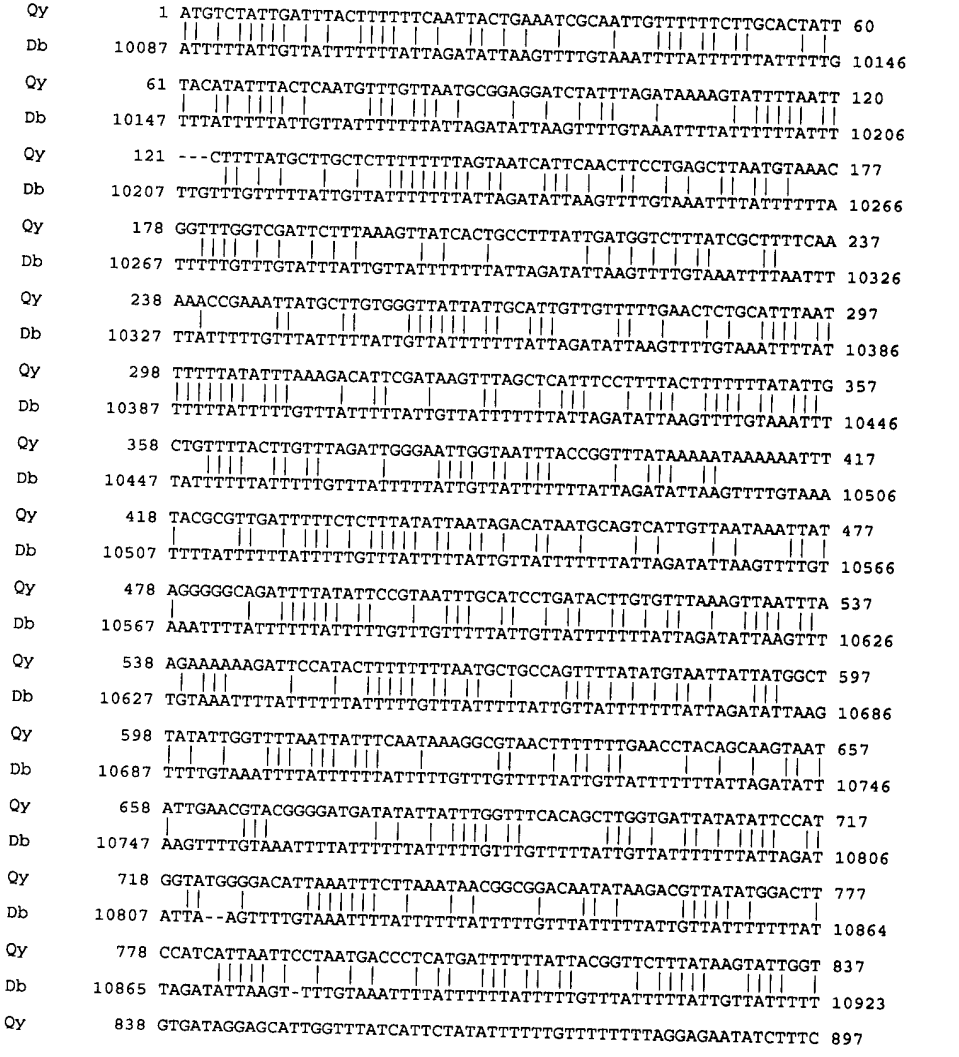
Qy         64 ATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTT 123
                | |||  |  | ||| ||          | | |||   |    | ||||  || ||
Db      48829 TTTTTTTTTTNTTTTTTTTTNNNTTNNTTTTTTTTTTTNTTTTTNNTTTTTTTTTTTTT 48888
```



RESULT 10
AX345086
LOCUS AX345086 15387 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 157 from Patent WO0200928.
ACCESSION AX345086
VERSION AX345086.1 GI:18492972
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1

AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 157 03-JAN-2002;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
source 1. .15387
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Query Match 11.1%; Score 116.6; DB 6; Length 15387;
Best Local Similarity 47.1%; Pred. No. 9.5e-07;
Matches 490; Conservative 0; Mismatches 544; Indels 7; Gaps 4;



```
Db      10924 TTATTAGATATTAAGTTTTGTAAATTTTATTTTTTATTTTTGTTTATTTTTATTGTTATT 10983

Qy        898 TTATTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTT 957

Db      10984 TTTTTATTAGATATTAAGTTTTGTAAATTTTATTTTTTATTTTTGTTTGTTTTTATTGTT 11043

Qy        958 GTGTTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACA 1017

Db      11044 ATTTTTTTATTTAGGTTTATTAATTTT-ATGTTTTTAGTTTGTTTTTAAGTTTTTTTTTT 11102

Qy       1018 GTTGGAGTTGTTTATGGATTT 1038

Db      11103 TTTTAAGTTTTTGATGTATTT 11123
```

RESULT 11
AC125567
LOCUS AC125567 170627 bp DNA linear HTG 09-NOV-2002
DEFINITION Rattus norvegicus clone CH230-9H22, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.
ACCESSION AC125567
VERSION AC125567.5 GI:24817949
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 170627)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 170627)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 170627)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 9, 2002 this sequence version replaced gi:23096537.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```
-------------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
    Center project name: GDDV
    Center clone name: CH230-9H22
-------------- Summary Statistics
    Assembly program: Phrap; version 0.990329
    Consensus quality: 162619 bases at least Q40
    Consensus quality: 164392 bases at least Q30
    Consensus quality: 165269 bases at least Q20
    Estimated insert size: 169231; sum-of-contigs estimation
    Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1   109158: contig of 109158 bp in length
*    109159   109258: gap of unknown length
*    109259   110280: contig of 1022 bp in length
*    110281   110380: gap of unknown length
*    110381   111540: contig of 1160 bp in length
*    111541   111640: gap of unknown length
*    111641   170627: contig of 58987 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .170627
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9H22"
     misc_feature    1. .1284
```

```
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    3636. .4924
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH341988"
     gap             109159. .109258
                     /estimated_length=unknown
     gap             110281. .110380
                     /estimated_length=unknown
     gap             111541. .111640
                     /estimated_length=unknown
ORIGIN
```

Query Match 11.1%; Score 116.6; DB 14; Length 170627;
Best Local Similarity 46.3%; Pred. No. 3.3e-07;
Matches 461; Conservative 0; Mismatches 524; Indels 10; Gaps 2;

```
Qy        20 TTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTACATATTTACTCAATGTT 79
Db     28097 TTTTTATTAAATTATTTTTTTATTTTATTTTTTTTTTTTATATTTTTTTTTATTATTTTTT 28156

Qy        80 TGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTTATGCTTGCTCTTTT 139
Db     28157 TTTTTTTTTTTATTATTAATATATTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTT 28216

Qy       140 TTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGTCGATTCTTTAAAGT 199
Db     28217 TTATTATATTTATTTTAATTTTTAAATTTTATTTTATATTTTTAATTTTTATTTTTTTTT 28276

Qy       200 TATCACTGCCTTTA--------TTGATGGTCTTTATCGCTTTTCAAAAACCGAAATTATG 251
Db     28277 TATATTTAAATTTTATTTATATTTTTTTTTTTTTATTATTTTTTAATTTTTTTTTTTTAT 28336

Qy       252 CTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTAAA 311
Db     28337 TTTTTTTATTTTTATTTATTTAATATTTATTTATAATATTATTAAATTTTTTTTTATTTT 28396

Qy       312 GACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACTTGTT 371
Db     28397 TTTTAATTATTTTTTTATTTTTTTTTTTTTTTATTTTTATTTTTATTTTTTAATTTTTT 28456

Qy       372 TAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGATTTT 431
Db     28457 TATTTTTTTTTTTTTATTTTATTTTTTTTTTTTTATAATTTTTTTATTTAAATTATTATTT 28516

Qy       432 TCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCAGATTTT 491
Db     28517 ATATTTTATTTTTTTATTTATATATTTTTATTTTTATTTTTTATAATTTTTTTTTATTTT 28576

Qy       492 ATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGATTCC 551
Db     28577 TTTTATATTAATATTTTTTTTATTAATTATTTTTTTATTTTATTTTTTTTTTTATTTTTA 28636

Qy       552 ATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGGTTTTAA 611
Db     28637 TTTTTTTTTTTAAATTTTTTATATTTTTTTTTTTTTTTATTTTTTTTTTTAAT--TTTAAA 28694

Qy       612 TTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTACGGG 671
Db     28695 TATTATTTTTTTTTTAATTTTTTTATTTTTTTTTTTTTTTTTTTATATTTTTATTTTTT 28754

Qy       672 GATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATGGGGACATT 731
Db     28755 TTTTTTTTTTTATATATTATATATTTTTATTTTTTTAATAATAATTTTTTTATAATAATT 28814

Qy       732 AAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATTAATTCC 791
Db     28815 AAATTTTTTTTATTTTATTTTAAATAATATTATTTAAAAATTTTATTATTATTATATTAA 28874

Qy       792 TAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATAGGAGCATT 851
```

```
Db     28875 AAATATTTTTTTTTTTTTTTTTATTTTTTTTATTATTTATTTAAATATTTTATTATTATA 28934

Qy       852 GGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTATTATATGAGAG 911
Db     28935 TTTTTTTTATTTATTATTATTTTTTTATTTAAATTTATATATTTATATTTATTAATTATTT 28994

Qy       912 AAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAATTTATAC 971
Db     28995 TTTTTATATTATTATAATTTTTTTTAATTTTTATTTTTTAATTTTTATTTAAAATTATTT 29054

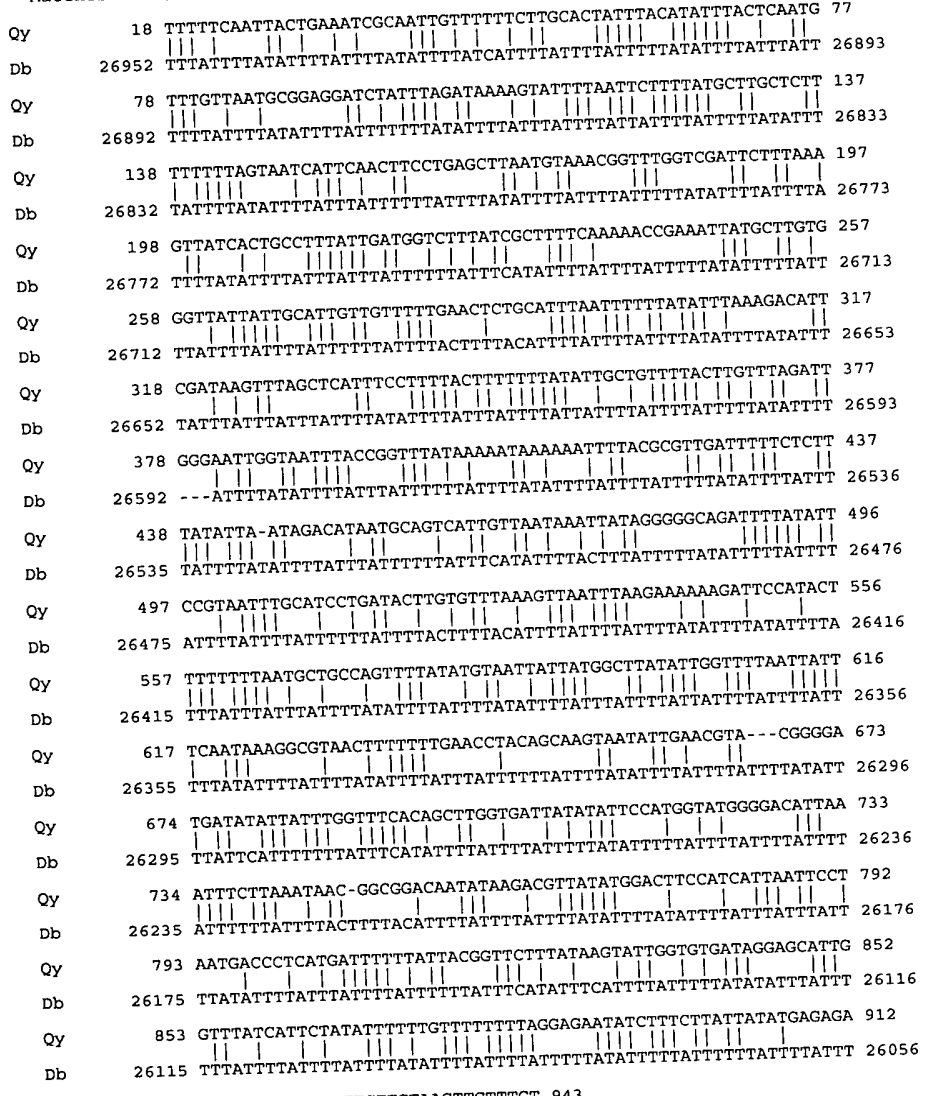
Qy       972 ATTAAACCCTTTTGATGCTTTTAATCGATTGATTT 1006
Db     29055 ATTATTTTATTTTTTTATTTTTATTTAATTTTTTT 29089
```

```
RESULT 12
CR936360/c
LOCUS       CR936360              154625 bp    DNA     linear   HTG 09-AUG-2005
DEFINITION  Homo sapiens chromosome X clone RP13-440N7, WORKING DRAFT SEQUENCE,
            2 unordered pieces.
ACCESSION   CR936360
VERSION     CR936360.11  GI:72060596
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 154625)
  AUTHORS   Whitehead,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 9, 2005 this sequence version replaced gi:71470739.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            -------------- Project Information
            Center project name: bB440N7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 154439 bases at least Q40
            Consensus quality: 154491 bases at least Q30
            Consensus quality: 154512 bases at least Q20
            Insert size: 154525; sum-of-contigs
            Insert size: 179913; 5.3% error; agarose-fp
            Quality coverage: 8.13x in Q20 bases; sum-of-contigs Quality
            coverage: 7.20x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   102638: contig of 102638 bp in length
            *  102639   102738: gap of 100 bp
            *  102739   154625: contig of 51887 bp in length.
FEATURES            Location/Qualifiers
     source         1. .154625
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="X"
                    /clone="RP13-440N7"
                    /clone_lib="RPCI-13.2"
```

misc_feature 1. .102638
/note="assembly_fragment:00149
fragment_chain:1"
misc_feature 102739. .154625
/note="assembly_fragment:02943
fragment_chain:1"

ORIGIN

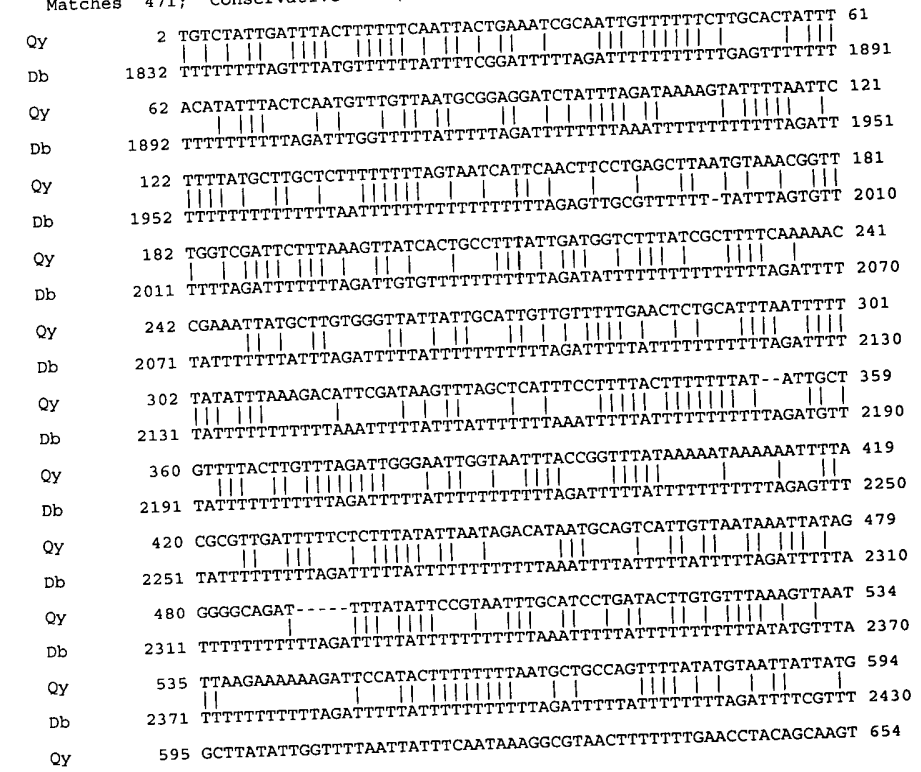
Query Match 10.8%; Score 113.4; DB 14; Length 154625;
Best Local Similarity 48.0%; Pred. No. 9e-07;
Matches 447; Conservative 0; Mismatches 476; Indels 8; Gaps 4;

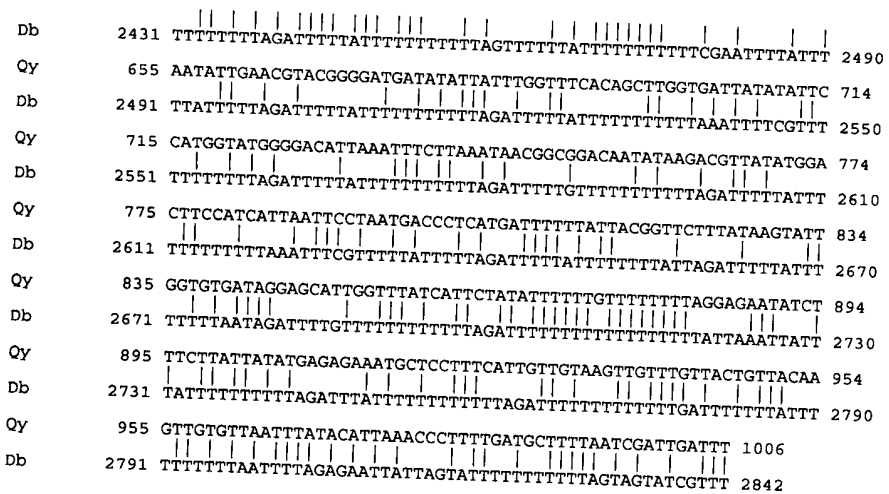


RESULT 13
AX251441
LOCUS AX251441 7442 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 409 from Patent WO0168912.
ACCESSION AX251441
VERSION AX251441.1 GI:15984864
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and oncogenes
JOURNAL Patent: WO 0168912-A 409 20-SEP-2001;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
source 1. .7442
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 10.8%; Score 113.2; DB 6; Length 7442;
Best Local Similarity 46.5%; Pred. No. 3.6e-06;
Matches 471; Conservative 0; Mismatches 533; Indels 8; Gaps 3;





```
Db       2431 TTTTTTTTAGATTTTATTTTTTTTTTTAGTTTTTATTTTTTTTTTTCGAATTTTATTT 2490
Qy        655 AATATTGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTC 714
Db       2491 TTATTTTTAGATTTTATTTTTTTTTTTAGATTTTATTTTTTTTTTAAATTTTCGTTT 2550
Qy        715 CATGGTATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGA 774
Db       2551 TTTTTTTTAGATTTTATTTTTTTTTTTAGATTTTGTTTTTTTTTTTAGATTTTTATTT 2610
Qy        775 CTTCCATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATT 834
Db       2611 TTTTTTTTAAATTTCGTTTTTATTTTTAGATTTTTATTTTTTTTATTAGATTTTTATTT 2670
Qy        835 GGTGTGATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCT 894
Db       2671 TTTTTAATAGATTTTGTTTTTTTTTTTTAGATTTTTTTTTTTTTTTTTTTATTAAATTATT 2730
Qy        895 TTCTTATTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAA 954
Db       2731 TATTTTTTTTTTAGATTTATTTTTTTTTTTTAGATTTTTTTTTTTTTGATTTTTTTATTT 2790
Qy        955 GTTGTGTTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTT 1006
Db       2791 TTTTTTTAATTTTAGAGAATTATTAGTATTTTTTTTTTTTAGTAGTATCGTTT 2842
```

RESULT 14
PFMAL1P3/c
LOCUS PFMAL1P3 67970 bp DNA linear INV 16-APR-2005
DEFINITION Plasmodium falciparum DNA from MAL1P3.
ACCESSION AL031746 AL844501
VERSION AL031746.9 GI:6594243
KEYWORDS .
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1
AUTHORS Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D., Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K., Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C., Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C., Cronin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J., Feltwell,T., Goble,A., Goodhead,I., Gwilliam,R., Hamlin,N., Hance,Z., Harper,D., Hauser,H., Hornsby,T., Holroyd,S., Horrocks,P., Humphray,S., Jagels,K., James,K.D., Johnson,D., Kerhornou,A., Knights,A., Konfortov,B., Kyes,S., Larke,N., Lawson,D., Lennard,N., Line,A., Maddison,M., McLean,J., Mooney,P., Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A., Rabbinowitsch,E., Rajandream,M.A., Rutter,S., Rutherford,K.M., Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R., Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L., Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and Barrell,B.G.
TITLE Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
JOURNAL Nature 419 (6906), 527-531 (2002)
PUBMED 12368867
REFERENCE 2
AUTHORS Murphy,L., Pain,A., Berriman,M., Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M., Rajandream,M., Hall,N. and Barrell,B.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT On Dec 16, 1999 this sequence version replaced gi:5763807.
For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES Location/Qualifiers
source 1. .67970
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="1"
gene complement(join(201. .257,455. .586,748. .804))
/gene="PFA0575c"
CDS /note="synonym: MAL1P3.01b"
complement(join(201. .257,455. .586,748. .804))
/gene="PFA0575c"
/note="Similar to Schizosaccharomyces pombe hypothetical protein c119.18 spbc119.18 SWALL:Q96VG1 (EMBL:AL022117) (69 aa) fasta scores: E(): 1.3e-06, 39.68% id in 63 aa, and to Homo sapiens protein 15e1.1 15e1.1 SWALL:15E1_HUMAN (SWALL:O43715) (76 aa) fasta scores: E(): 0.00046, 38% id in 50 aa, and to Mus musculus protein 15e1.1 15e1.1 SWALL:Q9D8Z2 (EMBL:AK007514) (76 aa) fasta scores: E(): 0.00054, 38% id in 50 aa"
/codon_start=1
/product="hypothetical protein, conserved"
/protein_id="CAD49075.1"
/db_xref="GI:23477009"
/db_xref="InterPro:IPR007918"
/db_xref="UniProt/TrEMBL:Q8I227"
/translation="MENRSKTISQNTIKAHVEANDECKEKKEKYLKCFNNWYKNNFLK
GDLTQACDDYYEDYQICVLNDLNKKGLGHLSNVEKEK"
gene complement(join(1392. .1461,1752. .2598,2748. .2848,
2990. .3276))
/gene="PFA0580c"
CDS /note="synonym: MAL1P3.01"
complement(join(1392. .1461,1752. .2598,2748. .2848,
2990. .3276))
/gene="PFA0580c"
/note="Originally annotated as Plasmodium falciparum; conserved hypothetical protein but pfam match and other evidences suggest that it is a putative deoxyribonuclease; earlier start site possible to make a 412 aa protein
Pfam match to entry PF01026 TatD_DNase, TatD related DNase, score 117.30, E-value 2.9e-31
Similar to Plasmodium falciparum conserved hypothetical protein, upf0006 family mal1p3.01 SWALL:Q9U0N6 (EMBL:AL031746) (412 aa) fasta scores: E(): 3.7e-150, 100% id in 412 aa, and to Saccharomyces cerevisiae putative deoxyribonuclease ybl055c ybl055c or ybl0512 or ybl0511 SWALL:YBF5_YEAST (SWALL:P34220) (418 aa) fasta scores: E(): 2.5e-19, 30.47% id in 397 aa"
/codon_start=1
/product="TatD-like deoxyribonuclease, putative"
/protein_id="CAD49076.1"
/db_xref="GI:23477010"
/db_xref="GOA:Q8I226"
/db_xref="InterPro:IPR001130"
/db_xref="UniProt/TrEMBL:Q8I226"
/translation="MKLVFHYIKYINVLFYISIIFLKSNSLKIYNDLRYISTVNKYKV
LQIKKRSNLKKNHNIRKMEDNESSFIDIGSNLTDKMFDGVYNSKKHENDLQNVLNRAK
NNNVDKIIITCTCLAEIDKSLKICETYDPEGKFLYLSAGVHPTNCYEFIDKNKHEEKE
IIAKKEYEEFIKYFKNEQVENSKMENGNKKICDGEKDMNNLNEILLEKNLDTIPGFKY
NEKDKEYLENLKNKIIKYPNRIVCIGEIGLDFDRLYFCSKYIQIKYFIFQLKLVQMFN
LPMFLHMRNCSETFFKIVDIYKFLFEKNGGVIHSFTDKEDIVHIIVQNYKNLYIGVNG
CSLKSLENINAVKKIPLNLLLLETDAPWCGVKKTHASYEYIKDTYEKRAYTNLKKIKN
IIKCDDNTIFKERNEPYNIADIAEITYKVREEAVPFDLFCKK"
gene 5005. .5496
/gene="MAL1P3.02"
CDS /note="synonym: PFA0585w"
5005. .5496
/gene="MAL1P3.02"
/note="contains possible signal sequence. Signal peptide predicted SignalP 2.0 HMM (Signal peptide probabilty 0.640, signal anchor probability 0.357) with cleavage site probability 0.504 between residues 27 and 28. ScanRegExp hit to PS00867, Carbamoyl-phosphate synthase subdomain signature 2"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB63557.1"

```
                     /db_xref="GI:6594245"
                     /db_xref="InterPro:IPR005479"
                     /db_xref="UniProt/TrEMBL:Q9U0N5"
                     /translation="MKLLNNRFVVLCPIIILFFFLNSVVLGNNNRNNINFHETENAAK
                     AMRKLLSGEINSIKLDNGDELKIKLNDEKHKDSTKWDKSYSFISNLEEEKYSQTDLFR
                     KKQEINEANTKIIEDRQEFYILNNDEIENIATRFVLENNFDELYIQSFKQSLIDIIQS
                     LNN"
misc_feature         8179. .10303
                     /note="Putative centromere"
gene                 14884. .20352
                     /gene="MAL1P3.03"
                     /note="synonym: PFA0590w"
CDS                  14884. .20352
                     /gene="MAL1P3.03"
                     /note="Putative ABC transporter
                     Pfam match to entry PF00005 ABC_tran, ABC transporter,
                     score 45.00, E-value 1.7e-09; HMMSmart hit to SM00382,
                     ATPases associated with a variety of cellular activities;
                     ScanRegExp hit to PS00211, ABC transporters family
                     signature
                     Similar to Schizosaccharomyces pombe ATP-binding cassette
                     transporter abc1 abc1 or spac9e9.12C SWALL:ABC1_SCHPO
                     (SWALL:Q92337) (1427 aa) fasta scores: E(): 4.2e-08,
                     21.46% id in 1146 aa"
                     /codon_start=1
                     /product="ABC transporter, putative"
                     /protein_id="CAB63558.1"
                     /db_xref="GI:6594246"
                     /db_xref="GOA:Q9U0N4"
                     /db_xref="InterPro:IPR001140"
                     /db_xref="InterPro:IPR003439"
                     /db_xref="InterPro:IPR003593"
                     /db_xref="InterPro:IPR011527"
                     /db_xref="UniProt/TrEMBL:Q9U0N4"
                     /translation="MTTYKENVGISNKGNKKKKSCQNISFLNFLSFDWIRPLINDLIK
                     GDIQELPNICRNFDVPYYASKLEENLRDIEVEDSEFYSEKNSSNEHVLHHCNSNDASE
                     KKVYNVYYHNILWSILKTFKFRIILIISFYILETLIVTLGGKFIDYYMRILEGQKIPV
                     YISFLKDFKVFSGLVVVMIMFFHLFFEALLHFYFHLFTINLKVSLMYFLYKINLCSNN
                     NHLQNPDAFYNTYRKFSSQTEIDEISRDFLSIGKNASSSSSGIKNNNKNIDNNKFVEN
                     DYIINFIKSTKKMEKDSLNENRSLPNVNIYNIMFSDVPSVTFFVTSCINLFNVFVKIF
                     MSFYVFHIKIGSNSVGIAIWLSIALYSAMILFEFLPSLFKSKYLIYRDKRIDNMHHVL
                     KEFKLIKMFNWESFAFKYINIFRMKEMKYCKIRLYLSNIGVFISSISSDIVEVVIFFI
                     YLKDRLNKKEEIKFTSIIMPLYVYKILISNVANFPNLVNNVMEGIVNIKRLNNYINDH
                     LYYNDIKNYFMYRTRYNEDYNIVVDKTFLQNENITSHDDGTSHNLKHLKNVIKNKLTN
                     MFKYFFFYHKMNYHKNIINKQILSGLLKNVDDNTNKKICFQEHKSNSTYNYNSSHIHE
                     KKEEYENIHNSSNSTMSNEFKEKKKNNEYIIKLENCSFGLSYDNKCDNDHILKNINFN
                     LKRNSLAIIIGNVGSGKSAFFHSILGDFNMTHGNLYIENFFKKMPILYVPQNSWLFMG
                     NIRSMILFGNEYNPLIYKYTILQSELLNDLSTIEHGDMKYINDDHNLSKGQKVRICLA
                     RALYEHYIHMHKLCTDYEKKLIQPNEILDKDLINNKNISSYNNKKSKLVNYNIPFNEN
                     YLQKCLMDDNNFYLYLLDDIFTSLDPSISKKIFSNLFCKEDNISFKDNCSFIISMNKS
                     TLDNFLIEDILDNVQYEVNIFEIQDKTLKYRGNISEYMEKNNLNITKESHWGYSNLNT
                     IDYTRIKLFDEVELNHVKHSNKMIYKEAYFVKGNTESVSFEIDSINKEYIKKMKKKNY
                     KKEHMNKNNKDNNNNNNNSNKDDHININMNDNHRNYNDINLGPNSTDDSPTVSSLGNE
                     YTLDTYTSNNSDKEEIVKPLYKDTHEEFNKSSSMPFVKSSSNMINNPSNFKYEDNSSS
                     FKGSISLETYLWYFQQVGFVLLTSVVIFMLISIFTDEIKFVFLTMMSIISKNNKEHSD
                     TILQKQVRYLEYFVILPIISLVTSGICFSMIIYGNITSAIKVHNNILYSILNAPLYIF
                     YNNNLGNIINRFIIDISAFDYGFLKRIYKAFFIFFRCILSSLLIIYMIRDCIFIFPFV
                     IILIYFFVFKRFSRGCKEAQRLYLSCHTPLCNIYSNALSGKNIINIYKKNTYHLDVYE
                     HYINNFRISYFFKWLINIWASLYIKIFILLLTTYIIMHPHLYASGIIKLYKEKNYVRI
                     LSTLGYCISFSARLGVIIKFLLCDYTHIEKEMCCVQRLEEFAKISNKENASMNKENEL
                     NVITTQTYKEKNENISDKISAIVEYKNVSLSSIINSSQDDESKKKYGIKFENVYVSYK
                     KKIPLVNGTYKYIDEEPSLKNINMYALKNQKIGIVGKSGAGKSTILLSILGLINISQG
                     KITVEGRDIRTYNRKGEDSIIGILAQSSFVFYNWNIRTFIDPYNNFTDDEIVHALKLN
                     GINLGKNDLYKYMHKQDMKSNYKKIIQTSKVINQSNDNTILLTNDCIRYLSLVRLYLN
                     RHKYKIILIDEIPIFNLNNSVHDELNSFLIGKAKSFNYIIRNHFPNNTVLIISHHANT
                     LSCCDYIYVLRKGEITYRCSYEDVKTQSELSHLLEMDD"
rRNA                 23896. .26044
                     /product="18s ribosomal RNA A-type"
                     /note="18s rRNA exppressed in asexual stage parasites"
misc_RNA             26045. .26438
                     /product="ITS1 A-type"
                     /note="ribosomal operon internal transcribed spacer 1
                     expressed in asexual stage parasites"
gene                 26439. .26559
                     /gene="mal1_5.8srRNA"
rRNA                 26439. .26559
                     /gene="mal1_5.8srRNA"
                     /product="5.8S ribosomal RNA"
                     /note="5.8S ribosomal exppressed in asexual stage
                     parasites
                     hit to vivax A-type 5.8S ribosomal percent id: 88.00 none"
misc_RNA             26560. .27435
                     /product="ITS2 A-type"
                     /note="ribosomal operon internal transcribed spacer 2

Query Match          10.6%;  Score 111.6;  DB 2;  Length 67970;
Best Local Similarity  46.7%;  Pred. No. 2.2e-06;
Matches  493;  Conservative    0;  Mismatches  554;  Indels    9;  Gaps    4;

Qy          1 ATGTCTATTGATTTACTTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              ||   || |   |||  || || ||||| |  |||     | |||| | |   | |||
Db       9284 ATAATTAATTTATTATATTATTTAATTAATTTAATAATTTTAATTTTATTTATTATTATA 9225

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAA--AAGTATTTTAA 118
              |  ||| || | |||  ||  | |          || |||||     |  || || |||||
Db       9224 TTTTTATATATTTAATTATTTATTAGCGTATATATTTATTTTATATATTAAATAATTTAA 9165

Qy        119 TTCTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACG 178
              ||  |||||          | | ||   | ||    | | ||  | |   |||| ||||
Db       9164 TTAATTTATTAAATTATTATATTAATTTTATATATTTAATTATTTAATATAATATAAATT 9105

Qy        179 GTTTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAA 238
              ||||  |  |||  | |  ||| | |  ||| |  | |  ||  | | |||
Db       9104 GTTTTATTTATTAAATTATTATATTAATATATTTTTATTTATTTAAAATATTATATTAAA 9045

Qy        239 AACCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATT 298
              | ||  |  |  |  |  ||  | | | ||| |  |       ||||||
Db       9044 TAATAAAGACAATATATTAAACATATAAATTAATATAATATTTTATTTTATTTATTAATT 8985

Qy        299 TTTTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGC 358
              | ||  || ||      ||| ||   |||  |  ||    | |  || ||||||  |
Db       8984 TGTTTAATATATTATTATTTTATTTAATTATTTATTTAATATATTATTATTTTATTTAAT 8925

Qy        359 TGTTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTT 418
              | ||| | |  |||    |  | || |||||||   | |||   |||| |  |||||
Db       8924 TATTTAATTATATTATTATTTTATTTATTAATTTA----TATATTTTAATATATTATTTT 8869

Qy        419 ACGCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATA 478
              |    || | |  | |||| ||| |  | |||||    |  | ||| | |||    |
Db       8868 ATTTATTTAATATAATATTTAATTTATTTAATATAATAATATTTTAATTATTTAATATAA 8809

Qy        479 GGGGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAA 538
              ||  |  |  ||| ||  |  |          || |  |  |  |  | ||||||
Db       8808 TATTTCATTTAATTCATTTAATATAATATTTTAATTATATTATTAATTTATATATTTTAA 8749

Qy        539 GAAAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTT 598
              |    |  |  || ||  | |  ||  |      |||||   | ||  |  |   ||
Db       8748 CATTTTATTTAATTTATTTAATATATATAATATTTTATTTATTAATTATATTATTAATTT 8689

Qy        599 ATATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATA 658
              ||||   ||  ||| | |  |          | |  | ||  |  ||       || ||
Db       8688 ATATATTTTAATATTTTATTTAATTTATTTAATATAATATTTTATTTATTAATTATATTA 8629

Qy        659 TTGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATG 718
              ||  |              || |||| ||||||  ||   |  |  |  |||| |  |  ||
Db       8628 TTAATTTATTTAATATAATATTTTATTTATTTAATTATATATATTATTAATTTATATATT 8569

Qy        719 GTATGGGGACATTAAATTTCTTAAATA-ACGGCGGACAATATAAGACGTTATATGGACTT 777
              ||       ||| ||||| || |||| |          ||| |  |  |  | ||| | ||
Db       8568 TTAATATTTTATTTAATTTATTTAATATAATATTTTATTTATTAATTATATTATTAATTT 8509

Qy        778 CCATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGT 837
              | ||  | |  |          | |  | | |||||    | | |||| |   |
Db       8508 ATTTAATATAATAGTTTATTTATTTAATTATATATATTATTAATTTATATATTTTAATAT 8449

Qy        838 GTGATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTC 897
```

```
Db       8448 TTTATTTAATTTATTTAATATAATATTTTATTTATTAATTATATTATTAATTTATATATT 8389

Qy        898 TTATTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTT 957

Db       8388 TTAATATTTTATTTAATTTATTTAATATATATAATATTTTATTTATTAATTATATTATTA 8329

Qy        958 GTGTTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACA 1017

Db       8328 ATTTATATATTTTAAT--ATTTTATTTAATTTATTTAATATAATATTTTATTTATTAATT 8271

Qy       1018 GTTGGAGTTGTTTATGGATTTGCAAAAATTAGATAA 1053

Db       8270 ATATTATTAATTTATATATTTTAATATTTTATTTAA 8235
```

RESULT 15
AC117342/c
LOCUS AC117342 175544 bp DNA linear HTG 15-NOV-2002
DEFINITION Rattus norvegicus clone CH230-372C13, WORKING DRAFT SEQUENCE, 2 unordered pieces.
ACCESSION AC117342
VERSION AC117342.5 GI:25013251
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 175544)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175544)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (10-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 175544)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 15, 2002 this sequence version replaced gi:23915286.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```
-------------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
    Center project name: GUDF
    Center clone name: CH230-372C13
-------------- Summary Statistics
    Assembly program: Phrap; version 0.990329
    Consensus quality: 168057 bases at least Q40
    Consensus quality: 169240 bases at least Q30
    Consensus quality: 170037 bases at least Q20
    Estimated insert size: 172468; sum-of-contigs estimation
    Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1   174007: contig of 174007 bp in length
*  174008   174107: gap of unknown length
*  174108   175544: contig of 1437 bp in length.
```

FEATURES Location/Qualifiers
source 1. .175544
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-372C13"
misc_feature 1. .2617
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 5845. .6767
/note="clone_boundary

```
                     clone_end:Sp6
                     site:
                     end_sequence:BZ217031"
     misc_feature    complement(98513. .99387)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ217027"
     gap             174008. .174107
                     /estimated_length=unknown
ORIGIN

  Query Match             10.5%;  Score 110.6;  DB 14;  Length 175544;
  Best Local Similarity   45.7%;  Pred. No. 2e-06;
  Matches  422;  Conservative    0;  Mismatches  500;  Indels    2;  Gaps    1;

Qy         45 TTTTTCTTGCACTATTTACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGA 104
              ||||| ||    | |||  || |||  ||  | ||  ||          | |||||
Db     135505 TTTTTTTTTTTTTTTTTGTATTTTTATCCTTTTTAATTTTATTATTTTTTTTATTTTAT 135446

Qy        105 TAAAAGTATTTTAATTCTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGA 164
              |     |||||  ||| | || |   ||  | | ||||||  | || |||     ||  |
Db     135445 TTTTTTTATTTATATTTTATTTTTATTTTTTTATTTTTTTTTTTATTATTTTTTTTATTTT 135386

Qy        165 GCTTAATGTAAACGGTTTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTT 224
                || || | ||    ||  |  | | | |||  ||||   |   | | || |   | ||
Db     135385 TTTTTATTTTAATTTATTTTTATAATTTATAATTTTATTTTTTTTTTATTTTTAAATTTTT 135326

Qy        225 TATCGCTTTTCAAAAACCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAA 284
              | |   | |     |      |      ||||  || |   ||  | ||   || || |  ||  |
Db     135325 TTTTAATAATATTTTATTATTTTTATTTTTTTTTTTTTATTTTTTTTTATTATAATTTTA 135266

Qy        285 CTCTGCATTTAATTTTTTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTAC 344
               | |   |||| ||||||||| |||       |||   ||  |||   |  |||  |||
Db     135265 TTATTATTTTATTTTTTTATTTTTTTTATTATTTATTATTTTTTATTTTTTTAATTTATT 135206

Qy        345 TTTTTTTATATTGCTGTTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAA 404
              || |||| | ||  |  |||  | | |   ||   ||||  || |||      ||| ||
Db     135205 TTATTTTTTTTTTTTTTAATTATATTTATTTTTTTTTTAATTTTTATTTTTTAATTTTTTAT 135146

Qy        405 AAATAAAAAATTTTACGCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATT 464
               ||      ||  ||    || |||||| | ||| | | | |      ||||  |   ||
Db     135145 TAAATTTTTATAATANTTTTTTATTTTTTTTTTTTTTTTAATTTTTATTAATATATAAATA 135086

Qy        465 GTTAATAAATTATAGGGGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTT 524
               |||||  |||           |  ||||  |||      || ||   |  |  |  |  | ||
Db     135085 TATAATATATTTATTTTATTATTTTTTTTATATATAAAATTTTTTATTATTTTTAATATT 135026

Qy        525 TAAAGTTAATTTAAGAAAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGT 584
              |     |||  ||||   | |      ||        |||||||  ||  |     ||||  |    |
Db     135025 TTT--TTATTTTATCATATTTTTTTATTATATTTTTTTTTTATTTTTATTTTTTTTTTAAT 134968

Qy        585 AATTATTATGGCTTATATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACC 644
               || |||    || | ||  ||||||||  |    ||| |    |||  || |  | ||
Db     134967 TTTTTTTAAATTTTTTTTTATTTTAATTTTAAATATATATTTATAATTTATAATTAATA 134908

Qy        645 TACAGCAAGTAATATTGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGA 704
              |      || |  | |  |   |        | || |      |    || |    ||    |
Db     134907 TTTTTTAATTTTTTTAAAGTTTTTTTTTTTTATTTTAATATATAATTAATTTTTTTAAAA 134848

Qy        705 TTATATATTCCATGGTATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGAC 764
               | ||  |  |          |   ||| |   || |       | | |
Db     134847 ATTAATTATATAATTATATTTTTTCTTTTTTTTTATTATTATTTTTTTTATTTTTTATT 134788

Qy        765 GTTATATGGACTTCCATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTT 824
               || | |    ||      ||||||  |  |        |  | ||| ||||  | |  |
Db     134787 TTTTTTTTATTTTTTTATTTTAATTTTTTTTATTTTAAAATAATTTATATTTTGTTATTA 134728

Qy        825 TATAAGTATTGGTGTGATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAG 884
              | | |    |  |     | ||  |||  ||| |  || |   |||||     |||  |
Db     134727 TTTTATATATATTAAATAAAGATTATTTTTTTTTTATTTTTTATTTTTTAAAGATTTAAAA 134668

Qy        885 GAGAATATCTTTCTTATTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTT 944
```

```
              |  ||||  ||| ||  ||     |  |     |  |||| || |||    |  ||
Db     134667 AAATTTATCAAACTTTTTTCATATATCATTTTTCTATAACATTTTTTTAAAGATTAGATT 134608

Qy         945 ACTGTTACAAGTTGTGTTAATTTA 968
                 ||   | || |  | ||| |
Db     134607 TTAATTTATAATTTTTGTTATTAA 134584


Search completed: May 10, 2006, 20:34:21
Job time : 3826.17 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:42:47 ; Search time 491.555 Seconds
(without alignments)
14276.972 Million cell updates/sec

Title: US-09-423-093B-1_COPY_9901_10953
Perfect score: 1053
Sequence: 1 atgtctattgatttactttt.....gatttgcaaaaattagataa 1053

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1053	100.0	14516	2	AAX06748	Aax06748 E. coli O
	2	1053	100.0	14516	3	AAZ56375	Aaz56375 Escherich
c	3	123.4	11.7	1300	12	ADP85917	Adp85917 Synthetic
c	4	120.6	11.5	1000	12	ADQ62833	Adq62833 Homopoly-
	5	120.6	11.5	1000	12	ADQ62832	Adq62832 Homopoly-
	6	116.6	11.1	15387	6	ABL32184	Abl32184 Human imm
	7	113.2	10.8	7442	4	AAS46686	Aas46686 Tumour su
	8	107.6	10.2	7814	4	AAS46530	Aas46530 Tumour su
	9	107.4	10.2	7597	6	ABL33013	Abl33013 Human imm
	10	107	10.2	6668	6	ABL33697	Abl33697 Human imm
c	11	105.6	10.0	6644	2	AAX33181	Aax33181 Base sequ
c	12	105.6	10.0	7372	2	AAX33182	Aax33182 Base sequ
c	13	105.6	10.0	7797	2	AAX33180	Aax33180 Cowpox vi
c	14	105.6	10.0	7996	2	AAX33184	Aax33184 Base sequ
	15	105.6	10.0	9539	4	AAS45347	Aas45347 Chemicall
	16	105.6	10.0	9539	6	ABK28180	Abk28180 DNA trans
	17	102.2	9.7	6292	4	AAS46735	Aas46735 Tumour su
	18	101.8	9.7	13511	6	ABL32281	Abl32281 Human imm
	19	101.6	9.6	5493	13	ADS89675	Ads89675 Oligonucl
	20	100.6	9.6	47108	6	ABK31511	Abk31511 Signal tr
	21	100.4	9.5	9539	4	AAS45346	Aas45346 Chemicall
	22	100.4	9.5	9539	6	ABK28179	Abk28179 DNA trans
	23	100.2	9.5	8392	6	ABL33490	Abl33490 Human imm
	24	99	9.4	14006	6	ABL33958	Abl33958 Human imm
	25	98	9.3	3683	8	ABZ10199	Abz10199 Haematopo
	26	98	9.3	73334	6	ABL34124	Abl34124 Human imm
	27	98	9.3	73334	6	ABL92318	Abl92318 Chemicall
	28	97.8	9.3	6898	6	ABN80222	Abn80222 Human che
	29	96.8	9.2	17967	6	ABL33015	Abl33015 Human imm
	30	96.2	9.1	6419	6	ABL32267	Abl32267 Human imm
c	31	95.6	9.1	1039	14	ACL64817	Acl64817 M. xanthu
c	32	95	9.0	612	14	ACL64894	Acl64894 M. xanthu
	33	94.2	8.9	5413	4	AAS46694	Aas46694 Tumour su
	34	94	8.9	3683	8	ABZ10053	Abz10053 Haematopo
	35	93.8	8.9	17594	6	ABL34026	Abl34026 Human imm
	36	93.4	8.9	6259	6	ABL32290	Abl32290 Human imm
	37	93.4	8.9	6286	4	AAS46591	Aas46591 Tumour su
	38	93.2	8.9	6767	4	AAS46608	Aas46608 Tumour su
	39	93.2	8.9	11222	10	ADB54318	Adb54318 Pretreate
	40	93.2	8.9	34548	6	ABL70604	Abl70604 Chemicall
	41	93	8.8	15548	6	ABL34155	Abl34155 Human imm
	42	92.6	8.8	5893	6	ABL32859	Abl32859 Human imm
	43	92.2	8.8	14950	6	ABL33257	Abl33257 Human imm
	44	91.8	8.7	83391	6	ABQ67094	Abq67094 Human ang
	45	91.6	8.7	11222	10	ADB54190	Adb54190 Pretreate

ALIGNMENTS

RESULT 1
AAX06748
ID AAX06748 standard; DNA; 14516 BP.
XX
AC AAX06748;
XX
DT 26-APR-1999 (first entry)
XX
DE E. coli O111 antigen gene cluster.
XX
KW O antigen; O111 antigen; diagnosis; wbdH gene; wzx gene; wzy gene;
KW wbdM gene; rfb gene; transferase; polymerase; flippase; diarrhoea;
KW haemorrhagic colitis; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 739. .1932
FT /*tag= a
FT /product= "galactosyl_transferase"
FT /note= "wbdH gene (ORF1), this region is specifically
FT claimed in Claim 6, and encodes AAW88300"
FT CDS 1983. .3101
FT /*tag= b
FT /note= "gmd gene (ORF2); encodes AAW88301"
FT CDS 3174. .3623
FT /*tag= c
FT /note= "wbdI gene (ORF3); encodes AAW88302"
FT CDS 3657. .5105
FT /*tag= d
FT /note= "manC gene (ORF4); encodes AAW88303"
FT CDS 5114. .6484
FT /*tag= e
FT /note= "manB gene (ORF5); encodes AAW88304"
FT CDS 6521. .7444
FT /*tag= f
FT /note= "wbdJ gene (ORF6); encodes AAW88305"
FT CDS 7441. .8607
FT /*tag= g
FT /note= "wbdK gene (ORF7); encodes AAW88306"
FT CDS 8646. .9911
FT /*tag= h

```
FT                   /product= "O_antigen_flippase"
FT                   /note= "wzx gene (ORF8), this region is specifically
FT                   claimed in Claim 6; encodes AAW88307"
FT   CDS             9901. .10953
FT                   /*tag= i
FT                   /product= "O_antigen_polymerase"
FT                   /note= "wzy gene (ORF9), this region is specifically
FT                   claimed in Claim 6; encodes AAW88308"
FT   CDS             10931. .11824
FT                   /*tag= j
FT                   /product= "colitose_or_glucose_transferase"
FT                   /note= "wbdL gene (ORF10); encodes AAW88309"
FT   CDS             11821. .12945
FT                   /*tag= k
FT                   /product= "colitose or glucose transferase"
FT                   /note= "wbdM gene (ORF11), this region is specifically
FT                   claimed in Claim 6; encodes AAW88310"
FT   CDS             13124. .14516
FT                   /*tag= l
FT                   /note= "gnd gene (partial sequence); encodes AAW88311"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;    98WO-AU000315.
XX
PR   01-MAY-1997;    97AU-00006545.
PR   22-JUL-1997;    97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88300, AAW88301, AAW88302, AAW88303, AAW88304, AAW88305,
DR   AAW88306, AAW88307, AAW88308, AAW88309, AAW88310, AAW88311.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 6; Fig 7; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Escherichia coli O111 O antigen. The gene cluster was
CC   isolated from a cosmid gene bank of E. coli M92 chromosomal DNA by
CC   screening with E. coli O111 antiserum. Open reading frames within the
CC   gene cluster were identified on the basis of homology to known sequences.
CC   The genes encode proteins (see AAW88300-11) involved the synthesis of
CC   sugars present in the polysaccharide antigen, and in the transport or
CC   processing of polysaccharide or oligosaccharide units. The use of nucleic
CC   acid molecules derived from particular assembly and transport genes,
CC   particularly wbd (transferase), wzx (flippase) and wzy (polymerase)
CC   genes, within O antigen gene clusters improves the specificity of methods
CC   for the detection and identification of O antigens, e.g. in testing food-
CC   or faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of enteropathogenic E. coli strains that cause
CC   diarrhoea and haemorrhagic colitis
XX
SQ   Sequence 14516 BP; 4582 A; 2221 C; 2971 G; 4742 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1053;  DB 2;  Length 14516;
  Best Local Similarity   100.0%;  Pred. No. 3e-169;
  Matches 1053;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9901 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 9960

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9961 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 10020

Qy        121 CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10021 CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 10080

Qy        181 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10081 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 10140

Qy        241 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10141 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 10200

Qy        301 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTTATATTGCTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10201 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTTATATTGCTG 10260

Qy        361 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10261 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 10320

Qy        421 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10321 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 10380

Qy        481 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10381 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 10440

Qy        541 AAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10441 AAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT 10500

Qy        601 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10501 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT 10560

Qy        661 GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10561 GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT 10620

Qy        721 ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10621 ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA 10680

Qy        781 TCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10681 TCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTG 10740

Qy        841 ATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10741 ATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTA 10800

Qy        901 TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10801 TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTG 10860

Qy        961 TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10861 TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT 10920

Qy       1021 GGAGTTGTTTATGGATTTGCAAAAATTAGATAA 1053
              |||||||||||||||||||||||||||||||||
Db      10921 GGAGTTGTTTATGGATTTGCAAAAATTAGATAA 10953

RESULT 2
AAZ56375
ID   AAZ56375 standard; DNA; 14516 BP.
XX
AC   AAZ56375;
XX
DT   17-MAR-2000  (first entry)
```

XX
DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:45.
XX
KW Flagellin; fliC; antigen; detection; ds.
XX
OS Escherichia coli.
XX
PN WO9961458-A1.
XX
PD 02-DEC-1999.
XX
PF 21-MAY-1999; 99WO-AU000385.
XX
PR 21-MAY-1998; 98AU-00003634.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Reeves PR, Wang L;
XX
DR WPI; 2000-072598/06.
XX
PT Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, feces, etc.
XX
PS Claim 3; Page 216-221; 245pp; English.
XX
CC AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all or
CC part of an Escherichia coli flagellin protein except a protein expressed
CC by E. coli H1, H7, H12 or H48 type strains. The present invention also
CC describes a method of detecting the presence of E. coli of a particular H
CC serotype in a sample, comprising specifically hybridising a nucleic acid,
CC preferably at least a pair, derived from a flagellating gene, specific
CC for a particular flagellin gene associated with the H serotype, to any
CC E.coli in the sample which contain the gene, and detecting any hybridised
CC molecules, identifying the presence of that serotype in the sample. (I)
CC are useful for: (1) detecting the presence of E. coli of H serotype in a
CC sample by hybridising at least one or a pair of (I) to any E. coli in the
CC sample and detecting the hybridised nucleic acid molecules; and (2) for
CC detecting the presence of both O and H-serotypes of E. coli by
CC hybridising at least one or a pair of (I) to any E. coli present in the
CC sample and detecting the hybridised nucleic acid molecules. (I) is
CC particularly useful for detecting the combination of O and H antigen.
CC Hybridised (I) when using at least one (I) is detected by southern blot
CC analysis and, when using a pair of (I), is detected by polymerase chain
CC reaction (PCR). AAZ56399 to AAZ56420 represent primers used in the
CC exemplification of the present invention
XX
SQ Sequence 14516 BP; 4582 A; 2221 C; 2971 G; 4742 T; 0 U; 0 Other;

Query Match 100.0%; Score 1053; DB 3; Length 14516;
Best Local Similarity 100.0%; Pred. No. 3e-169;
Matches 1053; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9901 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 9960

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9961 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 10020

Qy        121 CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10021 CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 10080

Qy        181 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10081 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 10140

Qy        241 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10141 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 10200

Qy        301 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10201 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTG 10260

Qy        361 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10261 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 10320

Qy        421 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10321 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 10380

Qy        481 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10381 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 10440

Qy        541 AAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10441 AAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT 10500

Qy        601 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10501 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT 10560

Qy        661 GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10561 GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT 10620

Qy        721 ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10621 ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA 10680

Qy        781 TCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10681 TCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTG 10740

Qy        841 ATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10741 ATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTA 10800

Qy        901 TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10801 TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTG 10860

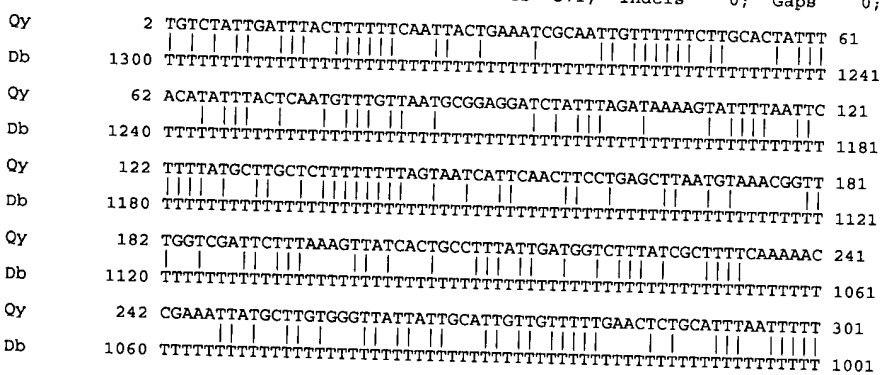
Qy        961 TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10861 TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT 10920

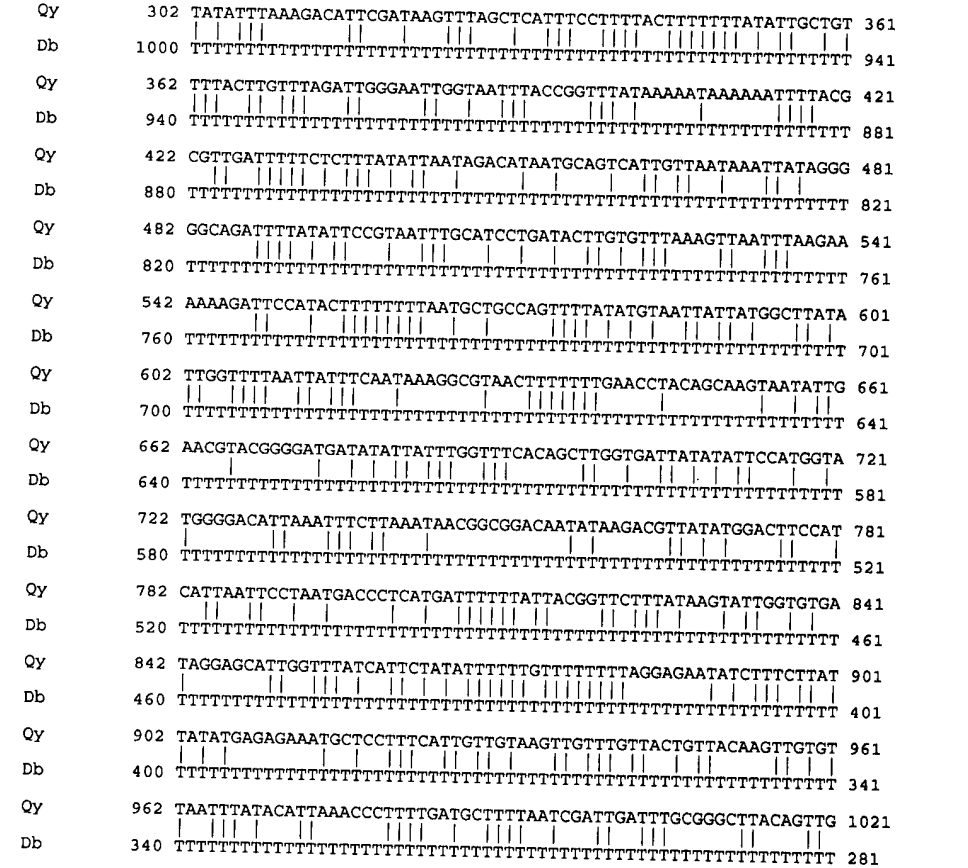
Qy       1021 GGAGTTGTTTATGGATTTGCAAAAATTAGATAA 1053
              ||||||||||||||||||||||||||||||||
Db      10921 GGAGTTGTTTATGGATTTGCAAAAATTAGATAA 10953
```

RESULT 3
ADP85917/c
ID ADP85917 standard; DNA; 1300 BP.
XX
AC ADP85917;
XX
DT 26-AUG-2004 (first entry)
XX
DE Synthetic construct #1.
XX
KW Nanopore data analysis ; polymer ; polynucleotide;
KW single-molecule resolution; ds.
XX
OS Synthetic.
XX
PN US2004110205-A1.
XX
PD 10-JUN-2004.
XX
PF 23-SEP-2003; 2003US-00668749.

XX
PR 23-SEP-2002; 2002US-0412959P.
XX
PA (WANG/) WANG H.
XX
PI Wang H;
XX
DR WPI; 2004-467656/44.
XX
PT Nanopore analysis system for analyzing polynucleotides includes nanopore
PT device having structure with aperture, and nanopore data analysis system
PT that can generate nanopore data points corresponding to each target and
PT non-target polymer.
XX
PS Disclosure; SEQ ID NO 1; 17pp; English.
XX
CC The invention relates to a nanopore analysis system that comprises a
CC nanopore device and a nanopore data analysis system. The nanopore device
CC has a structure having an aperture. The nanopore data analysis system can
CC generate nanopore data points corresponding to each target polymer and
CC each non-target polymer traversing the aperture; form a distribution
CC pattern of the data points; and analyse a distribution of target polymer
CC data points in the distribution pattern. The distribution pattern
CC includes data cluster(s). The nanopore data analysis system is stored on
CC a computer-readable medium. A mechanism is provided for analysing the
CC distribution of target polynucleotide data points in the distribution
CC pattern. The nanopore data analysis system can also analyse the
CC distribution of the non-target polynucleotide data points. It can also
CC determine a ratio between the target polynucleotide data points and the
CC non-target polynucleotide data points. It can further analyse the
CC distribution of target polynucleotide data points between the two data
CC clusters; compare the distribution of the target polynucleotide data
CC points between the two data clusters to a phosphorylation state standard
CC distribution; and determine a ratio of phosphorylated target
CC polynucleotide to non-phosphorylated target polynucleotides. It can also
CC determine a cluster score for the target polynucleotide data points in a
CC defined area; and compare the cluster score for the target polynucleotide
CC data points to a cluster score for a chemical integrity standard density
CC distribution for the defined area in a distribution of a target
CC polynucleotide standard. The invention is useful for analysing polymers,
CC such as polynucleotides. The inventive nanopore analysis system
CC potentially provides high speed sampling with single-molecule resolution,
CC which may enable unprecedented dynamic range and sensitivity in analysis
CC of samples containing charged polymers such as polynucleotides or
CC polypeptides. The present sequence is a synthetic construct used in the
CC invention.
XX
SQ Sequence 1300 BP; 1300 A; 0 C; 0 G; 0 T; 0 U; 0 Other;

Query Match 11.7%; Score 123.4; DB 12; Length 1300;
Best Local Similarity 44.9%; Pred. No. 5.9e-12;
Matches 466; Conservative 0; Mismatches 571; Indels 0; Gaps 0;

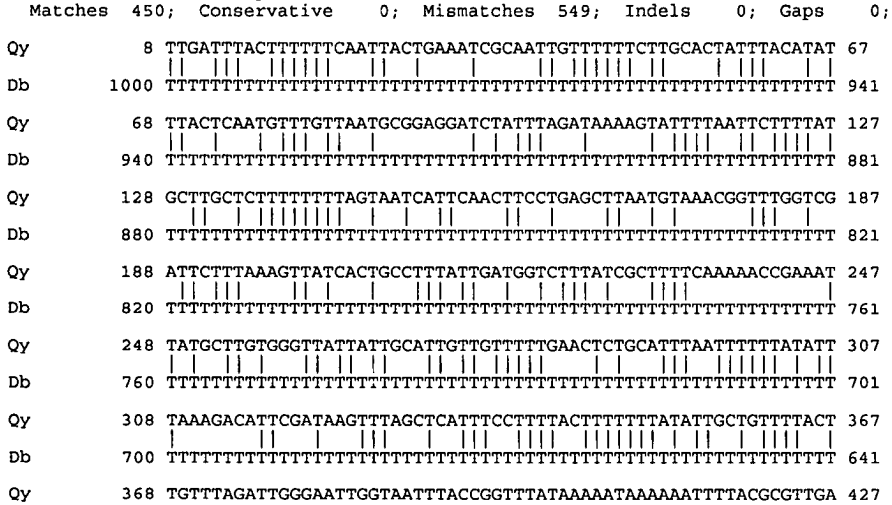


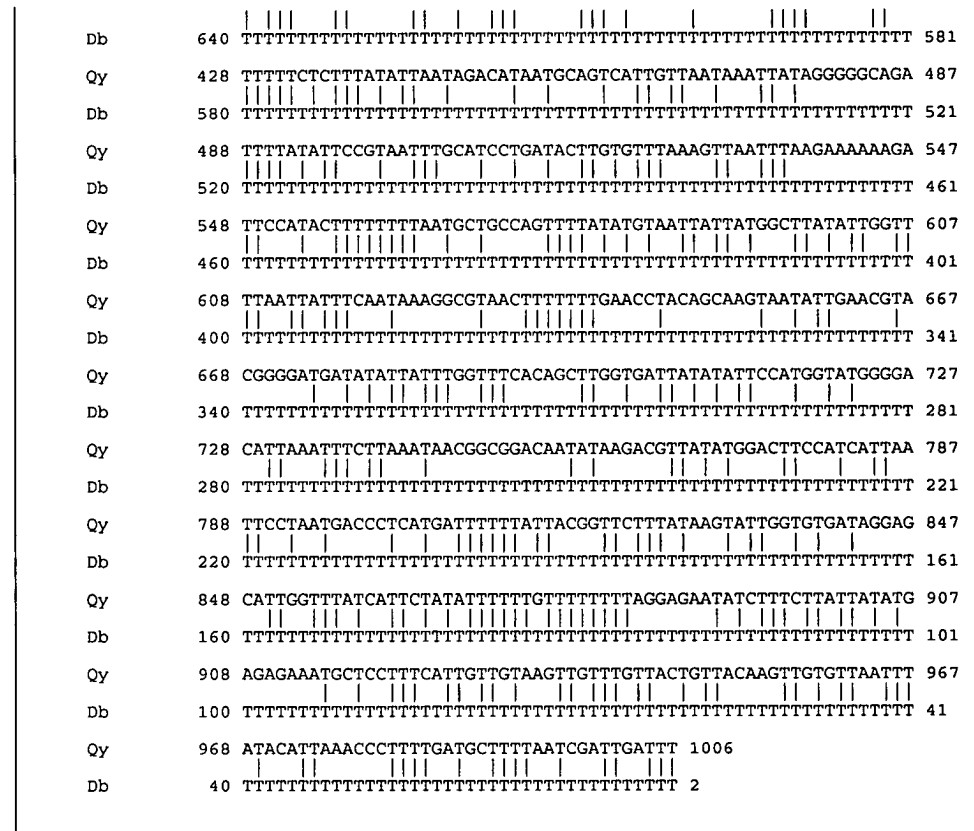


RESULT 4
ADQ62833/c
ID ADQ62833 standard; RNA; 1000 BP.
XX
AC ADQ62833;
XX
DT 07-OCT-2004 (first entry)
XX
DE Homopoly-A contaminant for RNaseH activity assay.
XX
KW ss; nuclease-mediated cleavage; target nucleic acid; RNaseH activity;
KW fluorophore; fluorescence quencher; fluorescent signal; HIV;
KW reverse transcriptase.
XX
OS Synthetic.
XX
PN WO2004059012-A1.
XX
PD 15-JUL-2004.
XX

PF 22-DEC-2003; 2003WO-US040879.
XX
PR 23-DEC-2002; 2002US-0436125P.
XX
PA (AMHP) WYETH.
XX
PI Olson MW, O'connell JF;
XX
DR WPI; 2004-543471/52.
XX
PT Detecting a nuclease-mediated cleavage of a target nucleic acid, useful
PT for detecting and monitoring RNase H activity, comprises hybridizing a
PT target nucleic acid to a fluorescently labeled oligonucleotide probe.
XX
PS Example 1; SEQ ID NO 4; 61pp; English.
XX
CC The invention relates to a method of detecting a nuclease-mediated
CC cleavage of a target nucleic acid or measuring a RNaseH activity of an
CC agent by hybridizing a target nucleic acid to a fluorescently labelled
CC oligonucleotide probe complementary to the target nucleic acid and
CC containing a fluorophore at one terminus and a quenching group at the
CC other terminus and contacting the probe-target hybrid with an agent
CC having nuclease activity. When the oligonucleotide probe is unhybridised
CC to the target nucleic acid, the probe adopts a conformation that places
CC the fluorophore and quencher in such proximity that the quencher quenches
CC the fluorescent signal of the fluorophore. Formation of the probe-target
CC hybrid causes sufficient separation of the fluorophore and quencher to
CC reduce quenching of the fluorescent signal of the fluorophore followed by
CC contacting the probe-target hybrid with an agent having nuclease activity
CC to selectively cleave the target nucleic acid and thus release the intact
CC probe. Detection of the release of the probe is by measuring a decrease
CC in the fluorescent signal of the fluorophore as compared to the signal of
CC the probe-target hybrid. The methods are useful for detecting and
CC monitoring RNaseH activity in real time. The methods are also useful for
CC monitoring enzymatic degradation of an RNA-DNA duplex. In an example of
CC the invention, the activity of the RNaseH from E. coli or HIV reverse
CC transcriptase is measure by the method of the invention. in order to test
CC the specificity of the RNaseH cleavage of the target-probe complex, a DNA
CC or RNA contaminant is added to the assay reaction. This sequence
CC corresponds to the single stranded homopolymeric polyA RNA contaminant
CC for testing this activity.
XX
SQ Sequence 1000 BP; 1000 A; 0 C; 0 G; 0 T; 0 U; 0 Other;

Query Match 11.5%; Score 120.6; DB 12; Length 1000;
Best Local Similarity 45.0%; Pred. No. 1.8e-11;
Matches 450; Conservative 0; Mismatches 549; Indels 0; Gaps 0;





RESULT 5
ADQ62832
ID ADQ62832 standard; RNA; 1000 BP.
XX
AC ADQ62832;
XX
DT 07-OCT-2004 (first entry)
XX
DE Homopoly-U contaminant for RNaseH activity assay.
XX
KW ss; nuclease-mediated cleavage; target nucleic acid; RNaseH activity;
KW fluorophore; fluorescence quencher; fluorescent signal; HIV;
KW reverse transcriptase.
XX
OS Synthetic.
XX
PN WO2004059012-A1.
XX
PD 15-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US040879.
XX
PR 23-DEC-2002; 2002US-0436125P.
XX
PA (AMHP) WYETH.
XX
PI Olson MW, O'connell JF;
XX
DR WPI; 2004-543471/52.

XX
PT Detecting a nuclease-mediated cleavage of a target nucleic acid, useful
PT for detecting and monitoring RNase H activity, comprises hybridizing a
PT target nucleic acid to a fluorescently labeled oligonucleotide probe.
XX
PS Example 1; SEQ ID NO 3; 61pp; English.
XX
CC The invention relates to a method of detecting a nuclease-mediated
CC cleavage of a target nucleic acid or measuring a RNaseH activity of an
CC agent by hybridizing a target nucleic acid to a fluorescently labelled
CC oligonucleotide probe complementary to the target nucleic acid and
CC containing a fluorophore at one terminus and a quenching group at the
CC other terminus and contacting the probe-target hybrid with an agent
CC having nuclease activity. When the oligonucleotide probe is unhybridised
CC to the target nucleic acid, the probe adopts a conformation that places
CC the fluorophore and quencher in such proximity that the quencher quenches
CC the fluorescent signal of the fluorophore. Formation of the probe-target
CC hybrid causes sufficient separation of the fluorophore and quencher to
CC reduce quenching of the fluorescent signal of the fluorophore followed by
CC contacting the probe-target hybrid with an agent having nuclease activity
CC to selectively cleave the target nucleic acid and thus release the intact
CC probe. Detection of the release of the probe is by measuring a decrease
CC in the fluorescent signal of the fluorophore as compared to the signal of
CC the probe-target hybrid. The methods are useful for detecting and
CC monitoring RNaseH activity in real time. The methods are also useful for
CC monitoring enzymatic degradation of an RNA-DNA duplex. In an example of
CC the invention, the activity of the RNaseH from E. coli or HIV reverse
CC transcriptase is measure by the method of the invention. in order to test
CC the specificity of the RNaseH cleavage of the target-probe complex, a DNA
CC or RNA contaminant is added to the assay reaction. This sequence
CC corresponds to the single stranded homopolymeric polyU RNA contaminant
CC for testing this activity.
XX
SQ Sequence 1000 BP; 0 A; 0 C; 0 G; 0 T; 1000 U; 0 Other;

Query Match 11.5%; Score 120.6; DB 12; Length 1000;
Best Local Similarity 0.0%; Pred. No. 1.8e-11;
Matches 0; Conservative 450; Mismatches 549; Indels 0; Gaps 0;

```
Qy          8 TTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTACATAT 67
              :: :::  :::::::  ::  :    :     :: :::::: ::    : :::  : :
Db          1 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 60

Qy         68 TTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTTAT 127
              ::  :   : ::: ::  :          : : :::   :     : :::: :: :::: :
Db         61 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 120

Qy        128 GCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGTCG 187
               ::  : ::::::::  :  :  ::    ::  :   ::  : :      :::  :
Db        121 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 180

Qy        188 ATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAAAT 247
              :: :::    :: :   :  ::: ::  :  : ::: :   ::::              :
Db        181 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 240

Qy        248 TATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATT 307
              : :  :: :   :: :: ::   :: :: :::::    : :   :::  :::::: : ::
Db        241 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 300

Qy        308 TAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACT 367
              :        ::   :  :::   :  :::  ::::  ::::::: : ::  : :::: :
Db        301 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 360

Qy        368 TGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGA 427
              : :::   ::      ::  :  :::     ::: :     :          ::::    ::
Db        361 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 420

Qy        428 TTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCAGA 487
              :::::  : ::: : ::  :     :  :    :  :: ::  :   :: :
Db        421 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 480

Qy        488 TTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGA 547
              :::: : ::   :  :::   :  :  :  :: : :::    ::  :::
```

```
Db        481 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 540

Qy        548 TTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGGTT 607
              ::   :  :::::::  :  :     :::: : : :  :: :: :   :: : ::  ::
Db        541 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 600

Qy        608 TTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTA 667
              :: :: :::   :       :   :::::::     :          : : ::     :
Db        601 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 660

Qy        668 CGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATGGGGA 727
                  :  : : :: :::  :::      ::  :  :: : : ::   :  : :
Db        661 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 720

Qy        728 CATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATTAA 787
               ::   ::: ::   :              : :       :: : : :    :   : ::
Db        721 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 780

Qy        788 TTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATAGGAG 847
              ::  :  :     :  :  :::::: ::    :: ::: :   : ::  : :  :
Db        781 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 840

Qy        848 CATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTATTATATG 907
               ::  ::: :  :: : : :::::: ::::::::       : : ::: :: :: : :
Db        841 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 900

Qy        908 AGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAATTT 967
                  :  :  :::  :: :: :   :: ::: ::   : ::      :: : ::  :::
Db        901 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 960

Qy        968 ATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTT 1006
              :   ::      ::::  :  ::::  :   ::  :::
Db        961 UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU 999
```

RESULT 6
ABL32184
ID ABL32184 standard; DNA; 15387 BP.
XX
AC ABL32184;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 157.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine

```
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 157; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 15387 BP; 3742 A; 300 C; 3619 G; 7726 T; 0 U; 0 Other;

  Query Match             11.1%;  Score 116.6;  DB 6;  Length 15387;
  Best Local Similarity   47.1%;  Pred. No. 8.2e-11;
  Matches  490;  Conservative    0;  Mismatches  544;  Indels    7;  Gaps    4;

Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              || | |||||  | |  |||| | |  || | |  |    |   ||| || ||    | |
Db      10087 ATTTTTATTGTTATTTTTTTATTAGATATTAAGTTTTGTAAATTTTATTTTTTATTTTTG 10146

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
              |  || ||||  |      ||| ||| |     |  |  | ||    |        | |||||  ||
Db      10147 TTTATTTTTATTGTTATTTTTTTTATTAGATATTAAGTTTTGTAAATTTTATTTTTTATTT 10206

Qy        121 ---CTTTTATGCTTGCTCTTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAAC 177
                 || | |    |  |  |||||||| ||   ||| |  ||  | |  || || |
Db      10207 TTGTTTGTTTTTATTGTTATTTTTTTATTAGATATTAAGTTTTGTAAATTTTATTTTTTA 10266

Qy        178 GGTTTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAA 237
              ||||  | |   | | |      | |    |          | | | | | ||     ||
Db      10267 TTTTTGTTTGTATTTATTGTTATTTTTTTATTAGATATTAAGTTTTGTAAATTTTAATTT 10326

Qy        238 AAACCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAAT 297
                |       ||    ||    |||||| ||  |||      ||  |   | |  |||| ||
Db      10327 TTATTTTTGTTTATTTTTATTGTTATTTTTTTATTAGATATTAAGTTTTGTAAATTTTAT 10386

Qy        298 TTTTTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTG 357
              |||||||  |||     |  ||    |     ||    | |||    | |||  |||| || |||
Db      10387 TTTTTATTTTTGTTTATTTTTATTGTTATTTTTTTATTAGATATTAAGTTTTGTAAATTT 10446

Qy        358 CTGTTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTT 417
                ||||  || ||      |    |||| || |||      |  |||  ||
Db      10447 TATTTTTTATTTTTGTTTATTTTTATTGTTATTTTTTTTATTAGATATTAAGTTTTGTAAA 10506

Qy        418 TACGCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTAT 477
              |     ||  | ||| | ||||| || || |  ||  |    | |    | |    || |
Db      10507 TTTTATTTTTTATTTTTGTTTATTTTTATTGTTATTTTTTTTATTAGATATTAAGTTTTGT 10566

Qy        478 AGGGGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTA 537
              |       |  |||||| ||    |  |||  ||  | ||  || | ||  |  |||| ||
Db      10567 AAATTTTATTTTTTATTTTTGTTTGTTTTTATTGTTATTTTTTTTATTAGATATTAAGTTT 10626

Qy        538 AGAAAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCT 597
              | |||      |   |  ||||| || ||  |    ||| | | | || |   |||
Db      10627 TGTAAATTTTATTTTTTATTTTTGTTTATTTTTATTGTTATTTTTTTTATTAGATATTAAG 10686

Qy        598 TATATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAAT 657
              | |  |   ||| ||| |||    |       ||    | || ||      |  |  |  || |
Db      10687 TTTTGTAAATTTTATTTTTTATTTTTGTTTGTTTTTATTGTTATTTTTTTTATTAGATATT 10746

Qy        658 ATTGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCAT 717
              |      |||         | |  | |||| ||       ||| |  || | ||||  ||
Db      10747 AAGTTTTGTAAATTTTATTTTTTATTTTTGTTTGTTTTTATTGTTATTTTTTTTATTAGAT 10806

Qy        718 GGTATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTT 777
              ||   |    |||||||  |   | |      |  || |     |||||  |    |
Db      10807 ATTA--AGTTTTGTAAATTTTATTTTTTATTTTTGTTTATTTTTATTGTTATTTTTTTAT 10864

Qy        778 CCATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGT 837
                  ||||| |  |  | |   | ||  ||| || ||     | |||||   ||||  |
Db      10865 TAGATATTAAGT-TTTGTAAATTTTATTTTTTATTTTTGTTTATTTTTATTGTTATTTTT 10923

Qy        838 GTGATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTC 897
              |  |||       | ||| | | | |   |||||  |||| ||||     ||   | |
Db      10924 TTATTAGATATTAAGTTTTGTAAATTTTATTTTTTATTTTTGTTTATTTTTATTGTTATT 10983

Qy        898 TTATTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTT 957
              || ||||  || |  ||   |  |    ||  | |    | ||||||  | ||   |||
Db      10984 TTTTTATTAGATATTAAGTTTTGTAAATTTTATTTTTTATTTTTGTTTGTTTTTATTGTT 11043

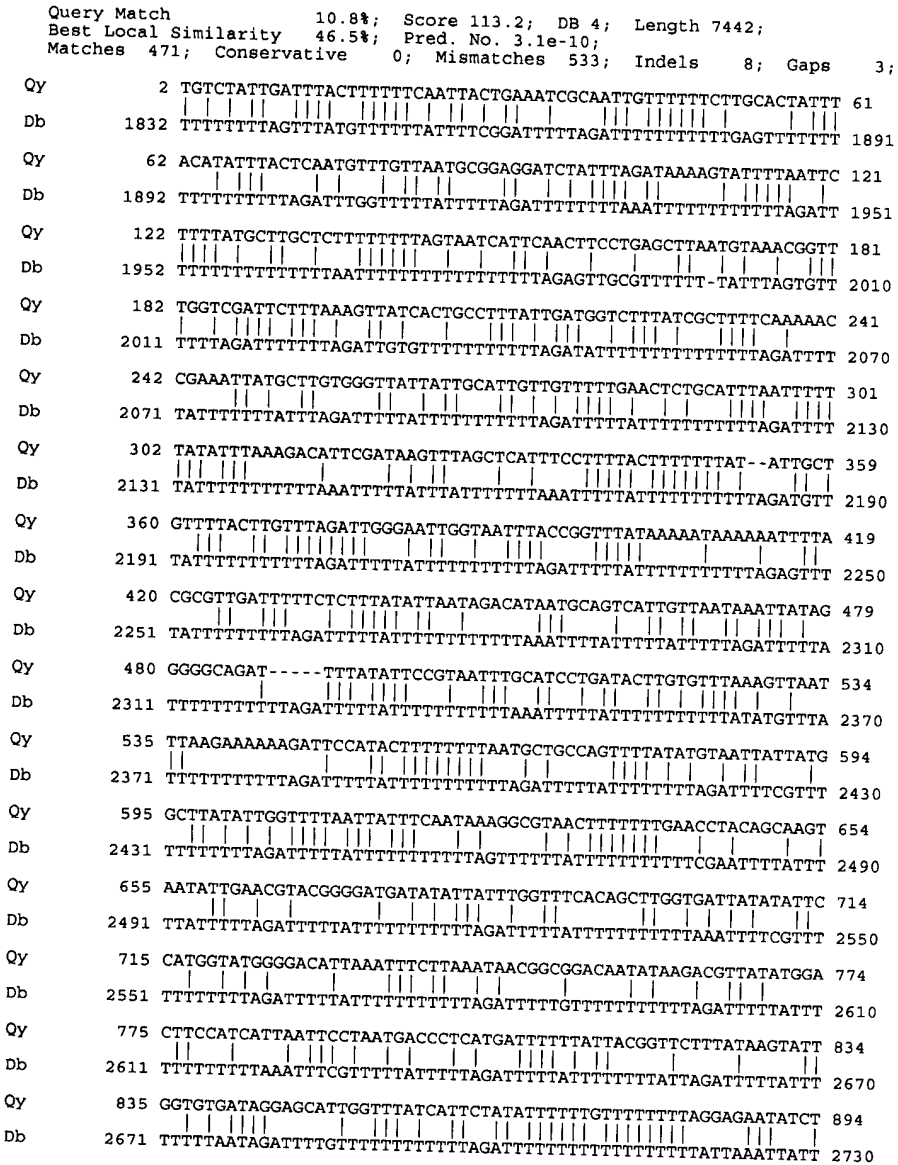
Qy        958 GTGTTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACA 1017
              | ||  | | ||  || |    |||| ||| ||||| |   ||  |  |    ||
Db      11044 ATTTTTTTATTTAGGTTTATTAATTTT-ATGTTTTTAGTTTGTTTTTAAGTTTTTTTTTT 11102

Qy       1018 GTTGGAGTTGTTTATGGATTT 1038
              ||  |||| || ||| ||||
Db      11103 TTTTAAGTTTTTGATGTATTT 11123


RESULT 7
AAS46686
ID   AAS46686 standard; DNA; 7442 BP.
XX
AC   AAS46686;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #409.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 409; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
```

CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 7442 BP; 1655 A; 214 C; 1934 G; 3636 T; 0 U; 3 Other;

Query Match 10.8%; Score 113.2; DB 4; Length 7442;
Best Local Similarity 46.5%; Pred. No. 3.1e-10;
Matches 471; Conservative 0; Mismatches 533; Indels 8; Gaps 3;





RESULT 8
AAS46530
ID AAS46530 standard; DNA; 7814 BP.
XX
AC AAS46530;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #252.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 252; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

```
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 7814 BP; 1677 A; 101 C; 1779 G; 4257 T; 0 U; 0 Other;

  Query Match           10.2%;  Score 107.6;  DB 4;  Length 7814;
  Best Local Similarity  45.3%;  Pred. No. 2.7e-09;
  Matches  469;  Conservative    0;  Mismatches  564;  Indels    3;  Gaps    2;

Qy         10 GATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTACATATTT 69
              | |||  |||||     |   |     | |    || | |||| | | | | |||  |||||
Db       3438 GGTTTTGTTTTTAGTTTAGGTTTTTTAGTTTTTTTATTTTTTGGTAATGTTTGAATATTA 3497

Qy         70 ACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTTATGC 129
              | | ||| ||| |   |    |   | | ||         | |||||    |  |||| |
Db       3498 ATTTAATTTTTTTGGTTTATAAATTTTTTTTGTTTATTTATTATTTATTTGTTTTTTTAT 3557

Qy        130 TTGCTCTTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGTCGAT 189
              |   |  ||| |||   | ||    |  | ||  | |   |    | | ||   | |  |  ||
Db       3558 GTATTTATTTGTTTGTTTATTTGTTTATTTAGTTATTTATTTATTAAGTTTATTATGTAT 3617

Qy        190 TCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGA--AAT 247
              |  || | ||  | | |   || |||| | || | | |   | || |          |||
Db       3618 TTATTTATGTGTTTATTTATTTAATTGTTCGTTTGTTTGTTTATTTATTTGTTAGTTAAT 3677

Qy        248 TATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATT 307
              |    |  |  |||||  |  |||  | | |||       |   ||  || |||  || |
Db       3678 TTATTTATTTATTTATTTGTTTATTTATTTGTTTATTTAATTATTTATATGTTTATTAAT 3737

Qy        308 TAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACT 367
              ||       ||   |    ||||  | |||   || || ||||  ||   ||  |  ||||
Db       3738 TAGTTCGTTTATTTGTTATTAGTTTATTAGTTTATAATTTTATTTGTTTATTTATTTATG 3797

Qy        368 TGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGA 427
                |||   |    | ||  | ||||      |||||  |  ||   | || |      |
Db       3798 AATTTGTTTATTTATTTATTTATTTGTTTATTTATTTATTTATTTATTTATTTATTTATT 3857

Qy        428 TTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCAGA 487
              | ||| | | | ||||  |  |  ||||  | | ||| || ||  ||| |          |
Db       3858 TATTTATTTATTTATTTGTTTA-TTAATTTATTTATTATTTATTTATTTTTTTATTTATT 3916

Qy        488 TTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGA 547
              | ||   ||   || |          |   |   ||   ||    | || |||||   |  |
Db       3917 TATTTATTTATTTATTAGTTTATTTATTTATTTATTTATTAATTTATTTATTTATTATTT 3976

Qy        548 TTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGGTT 607
              |        | ||| || ||       | || | ||| || |   ||  || || ||   |||
Db       3977 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTAGTTTCGGTTTTTAGGAGTT 4036

Qy        608 TTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTA 667
              |||      |   ||      |  |||| ||    |         |||          || ||
Db       4037 TTATGAGAATAGGTAGTATTTGATTTTTTGTTTTTGCGTTGTTTCGTATATGATAATGTT 4096

Qy        668 CGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATGGGGA 727
                ||  ||| |   ||| | | |      |  |   |  | | ||  |||||
Db       4097 TTTTAATTTTATTTATGTTGTTGTAAAGATATAATTTTATTTTTTTTTATGGTTGAATAG 4156

Qy        728 CATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATTAA 787
              |||  |||   || |||       | |||||| | |||||       |  ||   |
Db       4157 TATTTTATTATATATATATATAAATATAATATAATATGTTATTTATTTATTTATTCGTTG 4216

Qy        788 TTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATAGGAG 847
              |   ||      |     || | | ||     |  | |  ||| |||   | |     |
Db       4217 ATGGATATTTATGTTGATTTTGTATTTTTGTTATTATGAATAGTGTTGTGATAAATATAT 4276

Qy        848 CATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTATTATATG 907
              | ||   |||  || |     || ||  ||| |||||     ||| | | |  |||| |
Db       4277 AAATGTAGTATTTTTTTGGTATTATTTATTTATTTAGTTATTTATTTATTTATTTATTA 4336

Qy        908 AGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAATTT 967
                    |  ||   |  || |   ||  ||  |    |||    ||   ||| |
Db       4337 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTGTTTATTAA 4396

Qy        968 ATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTTGGAGTTG 1027
                |   |||     || |     | || ||  ||| ||||      |||    |  | ||
Db       4397 TTTAGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 4456

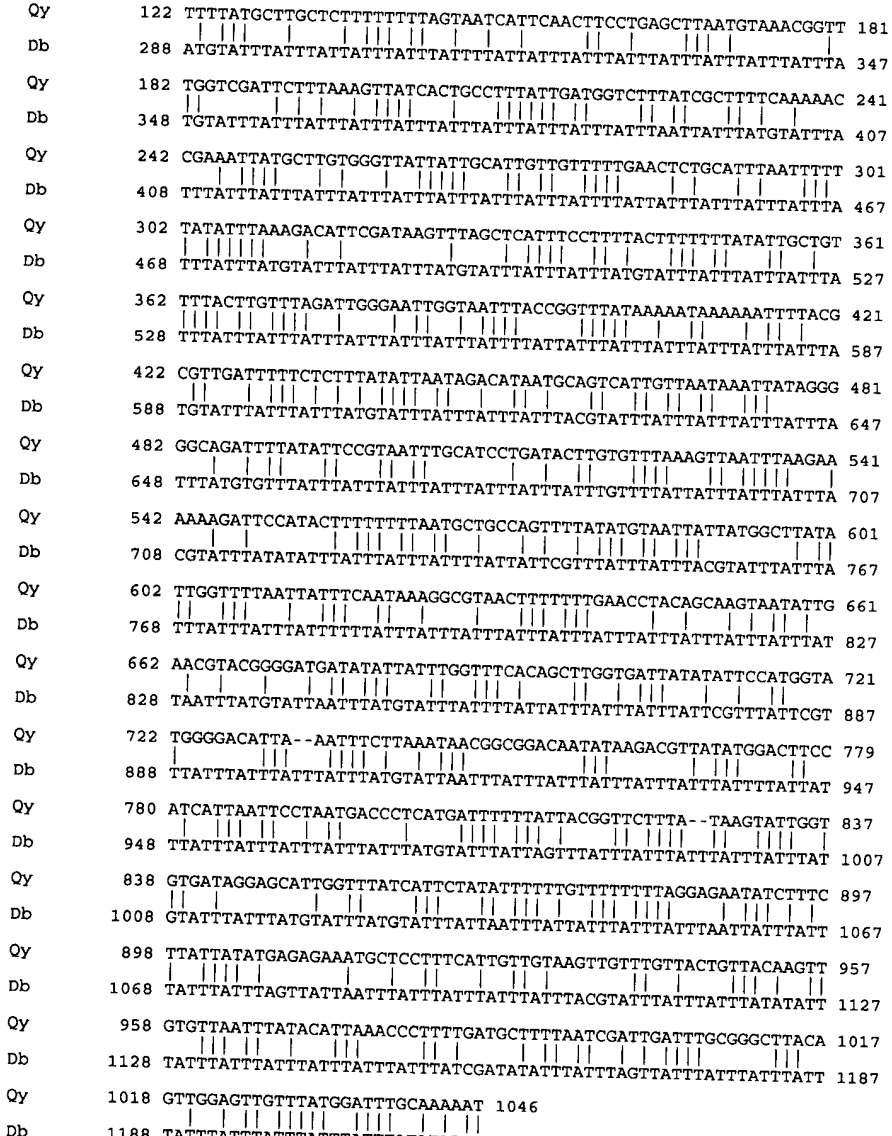
Qy       1028 TTTATGGATTTGCAAA 1043
                ||||  ||||  |||
Db       4457 TTTAGTTATTTATAAA 4472


RESULT 9
ABL33013
ID   ABL33013 standard; DNA; 7597 BP.
XX
AC   ABL33013;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 986.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 986; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 7597 BP; 1965 A; 109 C; 1790 G; 3733 T; 0 U; 0 Other;

  Query Match           10.2%;  Score 107.4;  DB 6;  Length 7597;
  Best Local Similarity  45.2%;  Pred. No. 3e-09;
  Matches  474;  Conservative    0;  Mismatches  571;  Indels    4;  Gaps    2;

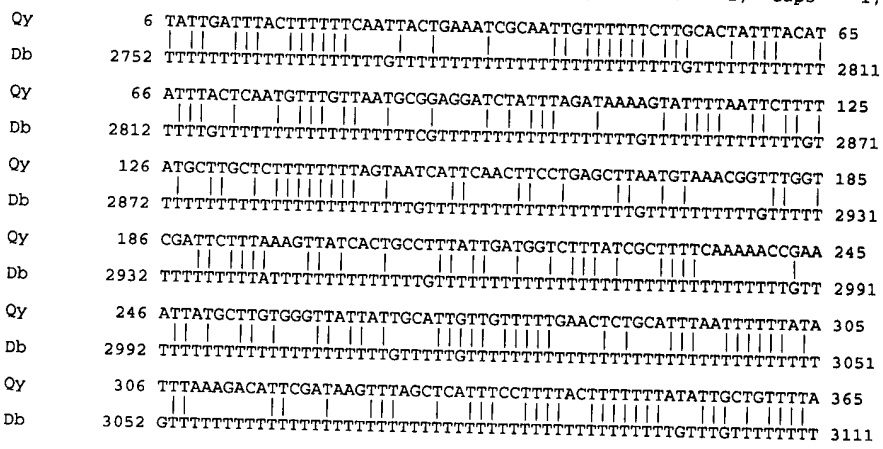
Qy          2 TGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTT 61
              | | | || ||||| || |||    |   |         | | |||| || ||    |||||
Db        168 TATTTGTTTATTTATTTATTTATTTATGTATTTATTTATTCGTTTATTATTTATTTATTT 227

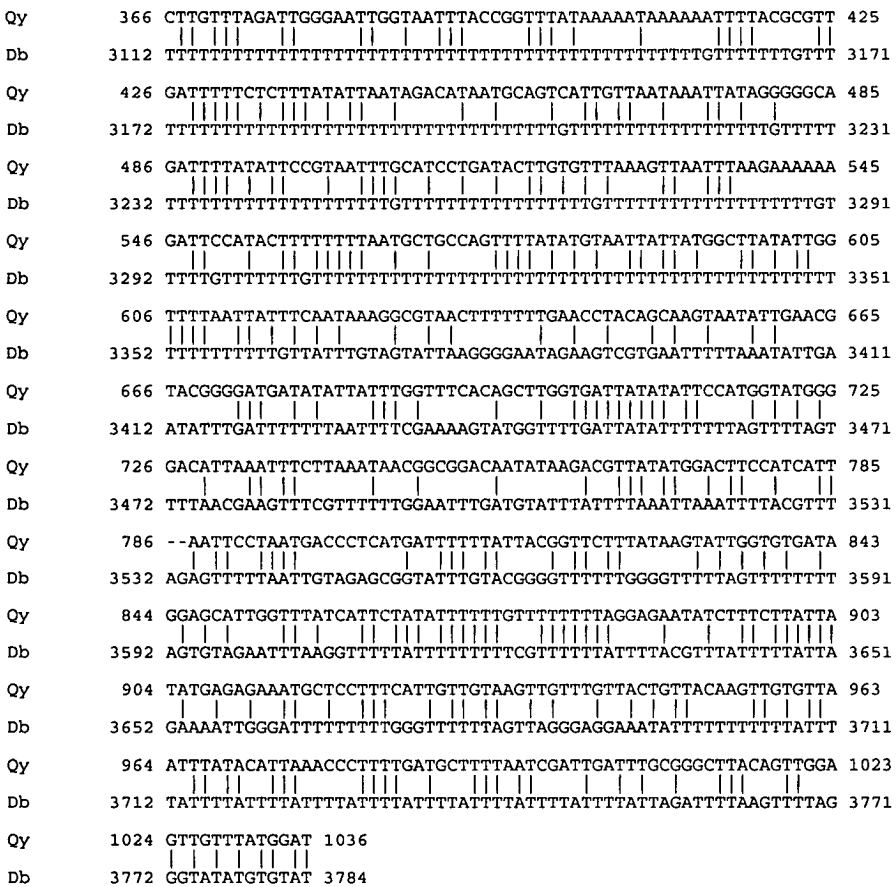
Qy         62 ACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTC 121
              |  ||||||   | |  ||  |           || ||||||  ||    | | || |||
Db        228 ATTTATTTATGAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 287
```



RESULT 10
ABL33697
ID ABL33697 standard; DNA; 6668 BP.
XX
AC ABL33697;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1670.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1670; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6668 BP; 1329 A; 328 C; 1736 G; 3274 T; 0 U; 1 Other;

Query Match 10.2%; Score 107; DB 6; Length 6668;
Best Local Similarity 44.6%; Pred. No. 3.5e-09;
Matches 461; Conservative 0; Mismatches 570; Indels 2; Gaps 1;

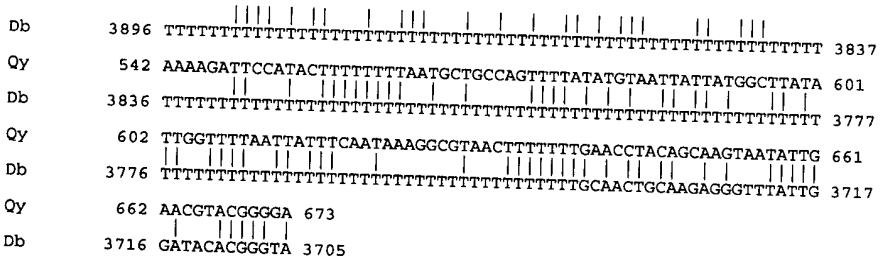




RESULT 11
AAX33181/c
ID AAX33181 standard; DNA; 6644 BP.
XX
AC AAX33181;
XX
DT 25-JUN-1999 (first entry)
XX
DE Base sequence of the plasmid pRx-ires-bsr.
XX
KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance; crmA;
KW bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
XX
OS Synthetic.
OS Cowpox virus.
XX
PN WO9913073-A2.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP004010.
XX
PR 08-SEP-1997; 97JP-00259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
PT New apoptosis-resistant virus-sensitive cell.
XX
PS Example 1; Page 38-41; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the plasmid
CC pRx-ires-bsr, which contains the cowpox virus bsr gene, and is used in an
CC example from the present invention
XX
SQ Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 U; 0 Other;

Query Match 10.0%; Score 105.6; DB 2; Length 6644;
Best Local Similarity 47.3%; Pred. No. 6e-09;
Matches 318; Conservative 0; Mismatches 354; Indels 0; Gaps 0;



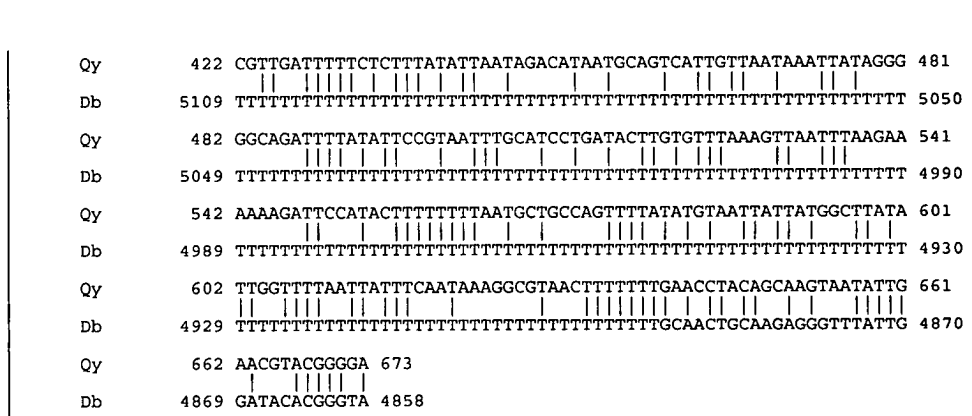


```
Db       3896 TTTT
```

PN WO9913073-A2.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP004010.
XX
PR 08-SEP-1997; 97JP-00259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
PT New apoptosis-resistant virus-sensitive cell.
XX
PS Example 1; Page 34-38; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the cowpox virus bsr gene which
CC is used in an example from the present invention
XX
SQ Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T; 0 U; 0 Other;

Query Match 10.0%; Score 105.6; DB 2; Length 7797;
Best Local Similarity 47.3%; Pred. No. 6e-09;
Matches 318; Conservative 0; Mismatches 354; Indels 0; Gaps 0;

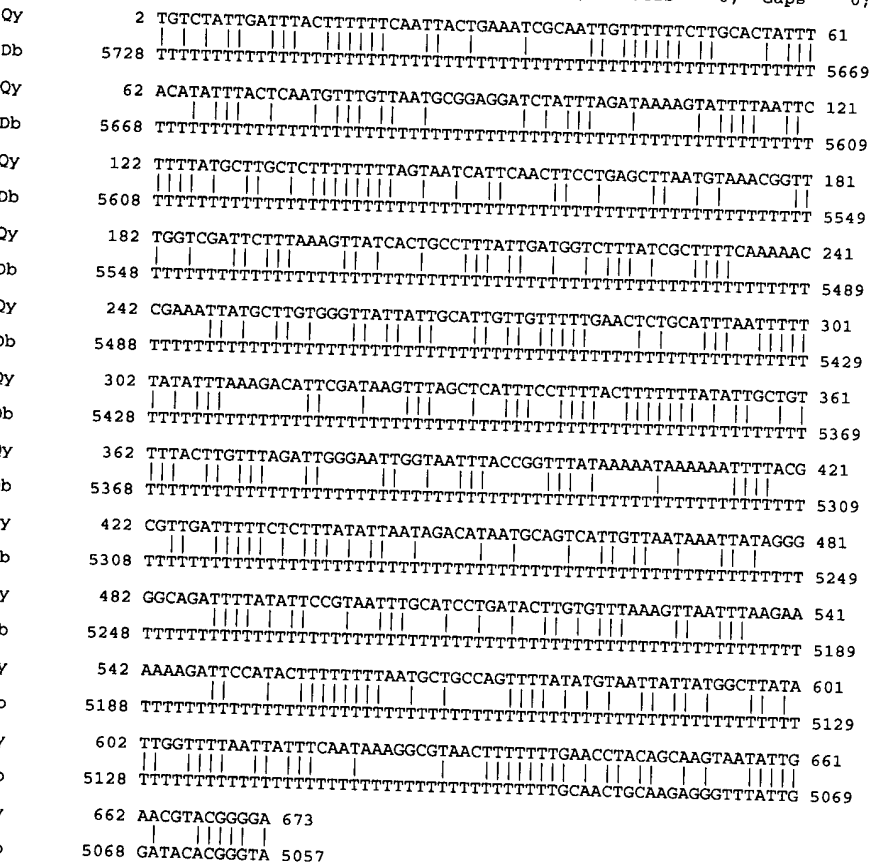




RESULT 14
AAX33184/c
ID AAX33184 standard; DNA; 7996 BP.
XX
AC AAX33184;
XX
DT 25-JUN-1999 (first entry)
XX
DE Base sequence of the plasmid pRx-Bcl 2-i-hCD 25.
XX
KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance; crmA;
KW bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9913073-A2.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP004010.
XX
PR 08-SEP-1997; 97JP-00259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
PT New apoptosis-resistant virus-sensitive cell.
XX
PS Example 3; Page 46-49; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated

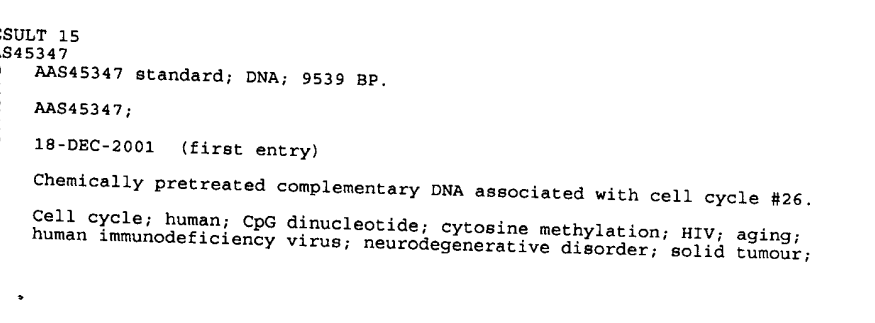
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the plasmid
CC pRx-Bcl 2-i-hCD 25, which contains the human Bcl-2 gene, and is used in
CC an example from the present invention
XX
SQ Sequence 7996 BP; 2463 A; 2015 C; 1829 G; 1689 T; 0 U; 0 Other;

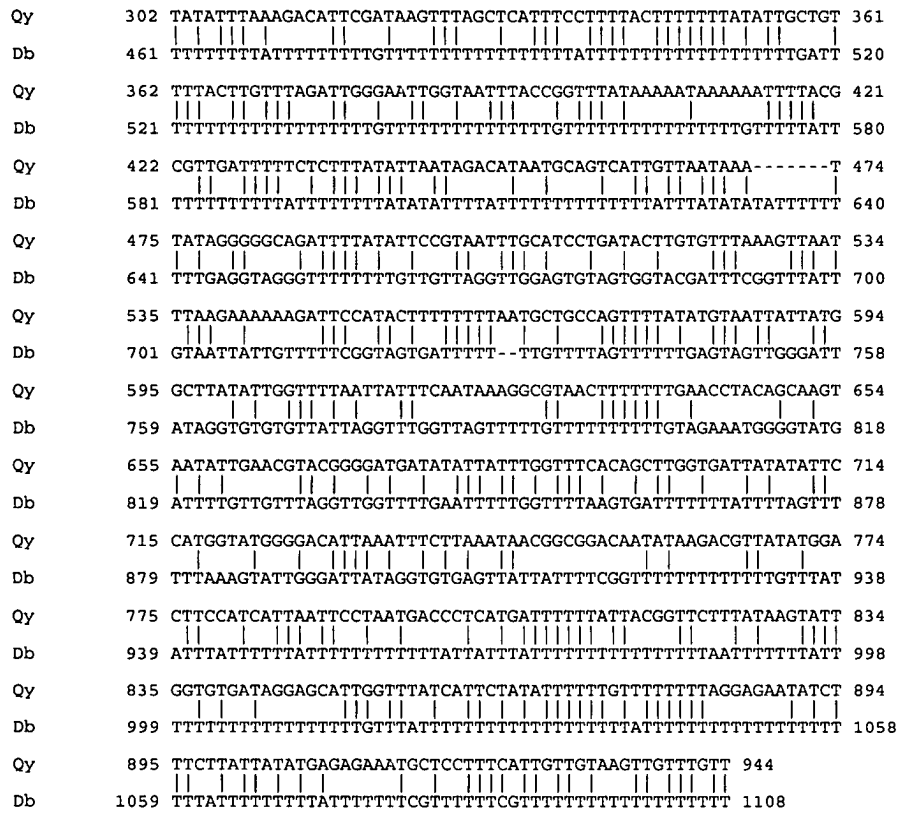
Query Match 10.0%; Score 105.6; DB 2; Length 7996;
Best Local Similarity 47.3%; Pred. No. 6e-09;
Matches 318; Conservative 0; Mismatches 354; Indels 0; Gaps 0;



RESULT 15
AAS45347
ID AAS45347 standard; DNA; 9539 BP.
XX
AC AAS45347;
XX
DT 18-DEC-2001 (first entry)
XX
DE Chemically pretreated complementary DNA associated with cell cycle #26.
XX
KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW PCR primer.
XX
OS Homo sapiens.
XX
PN WO200168911-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002945.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602751/68.
XX
PT Designing primers and probes for analyzing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle.
XX
PS Claim 1; SEQ ID NO 52; 28pp; English.
XX
CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers
XX
SQ Sequence 9539 BP; 2078 A; 121 C; 2281 G; 5059 T; 0 U; 0 Other;

Query Match 10.0%; Score 105.6; DB 4; Length 9539;
Best Local Similarity 46.0%; Pred. No. 6e-09;
Matches 437; Conservative 0; Mismatches 504; Indels 9; Gaps 2;





Search completed: May 10, 2006, 10:59:41
Job time : 494.555 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 11:00:02 ; Search time 3372.55 Seconds
(without alignments)
14608.170 Million cell updates/sec

Title: US-09-423-093B-1_COPY_9901_10953
Perfect score: 1053
Sequence: 1 atgtctattgatttactttt.....gatttgcaaaaattagataa 1053

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

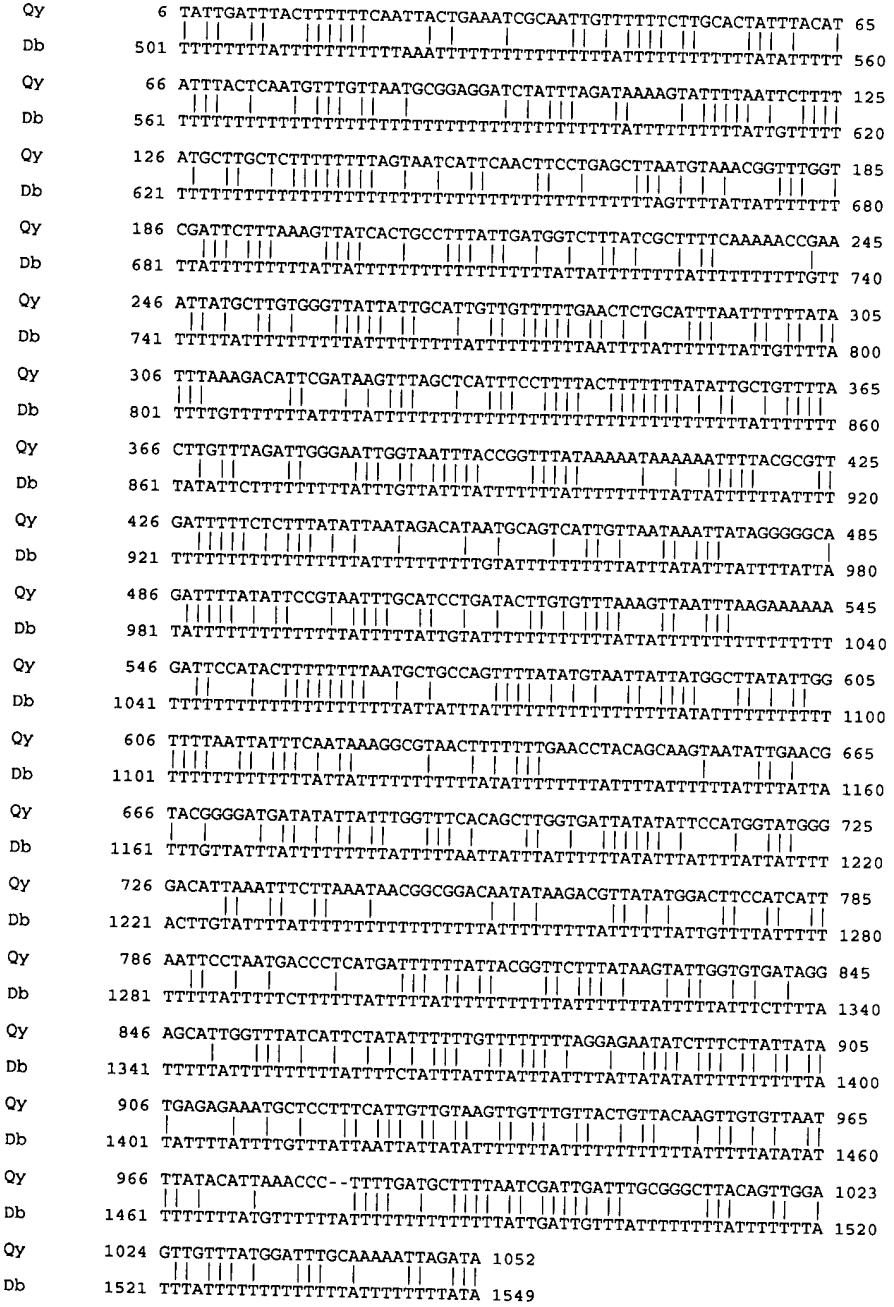
SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	142.2	13.5	1981	10	CL082000	CL082000 CH216-165
c	2	141	13.4	1330	8	DN705530	DN705530 CLJ60-A10
c	3	138.8	13.2	1299	8	DN705619	DN705619 CLJ60-F02
c	4	138.8	13.2	1626	6	CF238805	CF238805 AGENCOURT
c	5	137.8	13.1	1493	10	CL078589	CL078589 CH216-151
c	6	137.6	13.1	1536	10	CL078538	CL078538 CH216-151
	7	136.8	13.0	1434	10	AJ592058	AJ592058 Arabidops
c	8	136.8	13.0	1599	10	CL083840	CL083840 ISB1-2H14
	9	136.6	13.0	1507	10	AG346189	AG346189 Mus muscu
	10	136.2	12.9	1519	10	AG386893	AG386893 Mus muscu
	11	135.8	12.9	1334	10	AG341525	AG341525 Mus muscu
c	12	135.8	12.9	1594	10	CL038406	CL038406 CH216-46A
c	13	135.2	12.8	1367	10	CL082650	CL082650 CH216-168
c	14	134.8	12.8	1447	8	DN698671	DN698671 CLJ18-H11
	15	134.8	12.8	1482	10	AG332292	AG332292 Mus muscu
	16	134.6	12.8	1491	10	CG753221	CG753221 P048-2-A0
	17	134.6	12.8	1566	10	CG757757	CG757757 P053-1-D0
c	18	134.6	12.8	1721	8	DR145459	DR145459 49056454
c	19	133.8	12.7	1616	10	CL081995	CL081995 CH216-165
c	20	133.6	12.7	1632	10	CL082569	CL082569 CH216-167
	21	133.2	12.6	1453	10	AJ591978	AJ591978 Arabidops
c	22	132.6	12.6	1263	10	CL079431	CL079431 CH216-155
c	23	132.2	12.6	1641	10	CL078595	CL078595 CH216-151
	24	132	12.5	1288	10	CG744915	CG744915 P037-3-F0
	25	132	12.5	1433	10	CG745119	CG745119 P037-4-G0
c	26	132	12.5	1459	10	CL040025	CL040025 CH216-49A
c	27	132	12.5	1594	10	CL078613	CL078613 CH216-151
	28	131.8	12.5	1298	10	CG756607	CG756607 P051-4-C0
	29	131.6	12.5	1811	10	CG753732	CG753732 P048-4-G0
c	30	131.4	12.5	1531	10	CG748014	CG748014 P041-4-B0
	31	131.2	12.5	1313	10	CG751144	CG751144 P045-3-E0
	32	131.2	12.5	2157	10	CL081966	CL081966 CH216-165
	33	130.8	12.4	1269	10	CG757211	CG757211 P052-2-G0
	34	130.6	12.4	1353	10	CG744812	CG744812 P037-3-B0
	35	130.4	12.4	1373	10	CG750869	CG750869 P045-2-E1
	36	130.2	12.4	1407	10	AJ592026	AJ592026 Arabidops
	37	130	12.3	1317	1	AJ926430	AJ926430 AJ926430
	38	129.8	12.3	1228	10	AJ592059	AJ592059 Arabidops
	39	129.8	12.3	1276	10	CG754010	CG754010 P049-2-C0
	40	129.8	12.3	1376	10	CG747831	CG747831 P041-3-B0
c	41	129.8	12.3	1533	10	CL081968	CL081968 CH216-165
c	42	129.6	12.3	1364	8	DN693282	DN693282 CGX87-B03
	43	129.6	12.3	1559	10	CL646206	CL646206 CH213-110
	44	129.4	12.3	1324	10	CG746828	CG746828 P040-1-G1
	45	129.4	12.3	1594	10	CL110653	CL110653 ISB1-53P2

ALIGNMENTS

RESULT 1
CL082000
LOCUS CL082000 1981 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-165P18_RM4.1 CH216 Xenopus tropicalis genomic clone
CH216-165P18, genomic survey sequence.
ACCESSION CL082000
VERSION CL082000.1 GI:40537913
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1981)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM4 ctcaagggcatcggtcgagc
Class: BAC ends
High quality sequence start: 265
High quality sequence stop: 497.
FEATURES Location/Qualifiers
source 1. .1981
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-165P18"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"
ORIGIN

Query Match 13.5%; Score 142.2; DB 10; Length 1981;
Best Local Similarity 46.6%; Pred. No. 3.9e-15;
Matches 489; Conservative 0; Mismatches 558; Indels 2; Gaps 1;

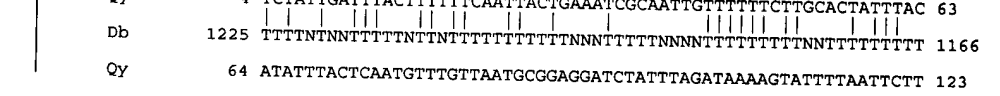


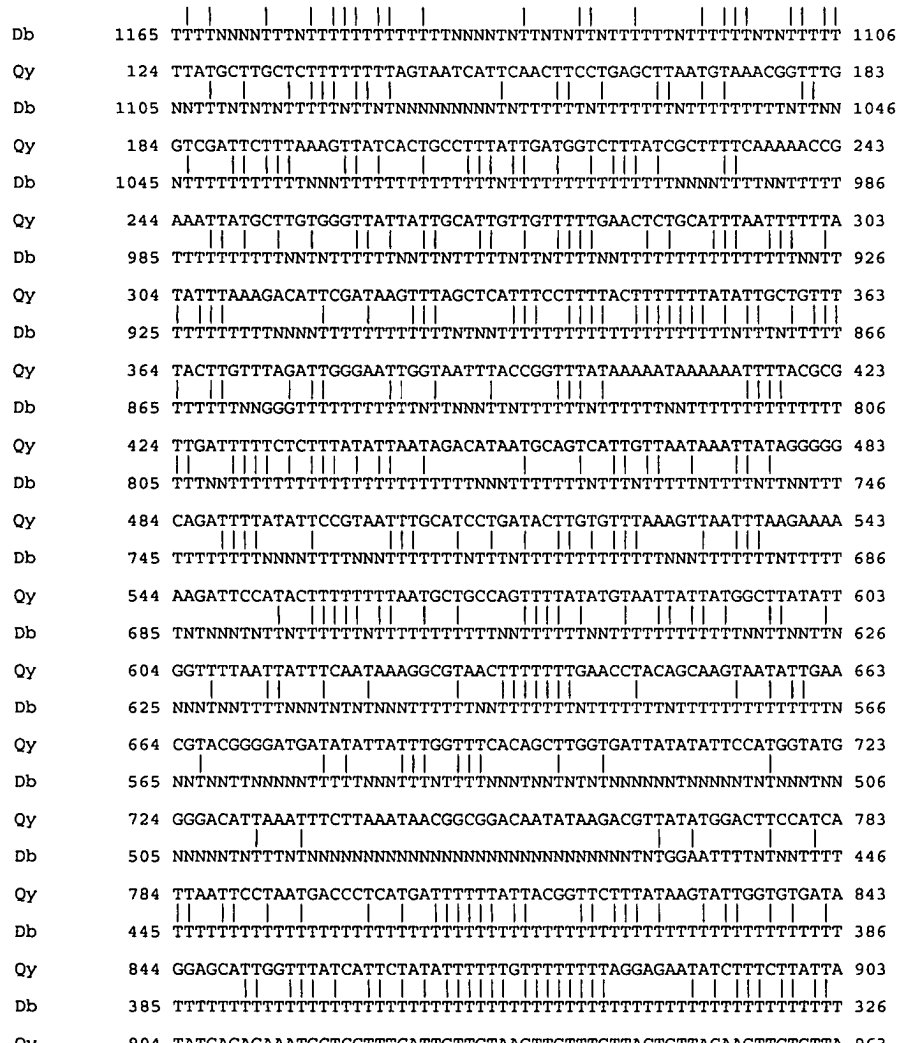
RESULT 2
DN705530/c
LOCUS DN705530 1330 bp mRNA linear EST 30-MAR-2005
DEFINITION CLJ60-A10.x1d-t SHGC-CLJ Gasterosteus aculeatus cDNA clone CLJ60-A10 3', mRNA sequence.
ACCESSION DN705530
VERSION DN705530.1 GI:62070570
KEYWORDS EST.
SOURCE Gasterosteus aculeatus (three spined stickleback)
ORGANISM Gasterosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes; Gasterosteidae; Gasterosteus.
REFERENCE 1 (bases 1 to 1330)
AUTHORS Kingsley,D.M., Peichel,C., Balabahdra,S., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
TITLE Expressed sequence tags from Gasterosteus aculeatus
JOURNAL Unpublished (2003)
COMMENT Contact: Grimwood, Jane
Stanford Human Genome Center
Stanford University School of Medicine
975 S California Ave, Palo Alto, CA 94304, USA
Tel: 650 320 5917
Fax: 650 320 5801
Email: jane@shgc.stanford.edu
Plate: 60
High quality sequence start: 189
High quality sequence stop: 449.

FEATURES Location/Qualifiers
source 1. .1330
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Bitrufjordur marine sticklebacks, Iceland"
/db_xref="taxon:69293"
/clone="CLJ60-A10"
/sex="mixed male and female"
/tissue_type="whole larva"
/dev_stage="21 day old larvae collected at Swarup Stage 30 (J. Embryol. Exp. Morphol 6: 373-383.1958)"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="SHGC-CLJ"
/note="Vector: Express 1; Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by Open Biosystems. First strand cDNA synthesis was primed with an 54 bp linker primer containing an oligodT sequence preceeded by a synthetic NotI site (first strand primer: 5'-GACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3'). Following second strand synthesis, cDNAs were made blunt at the end corresponding to the original 5 prime end of mRNA, and cloned directionally into the NotI and EcoRV sites of Express 1. Note that the EcoRV site is typically destroyed in the blunt end cloning, leaving a junction of the form 'xxxATC' (where is ATC is the second half of the EcoRV site, and xxx is derived from the cDNA sequence). A map of the Express 1 vector is available at: http://www.openbiosystems.com/cdna_library_construction_faq.php#8 The primary library was transformed and amplified in DH10B (T1 phage resistant) bacteria. Clones available from Open Biosystems: http://www.openbiosystems.com/stickleback"

ORIGIN

Query Match 13.4%; Score 141; DB 8; Length 1330;
Best Local Similarity 37.7%; Pred. No. 6.4e-15;
Matches 378; Conservative 0; Mismatches 625; Indels 0; Gaps 0;





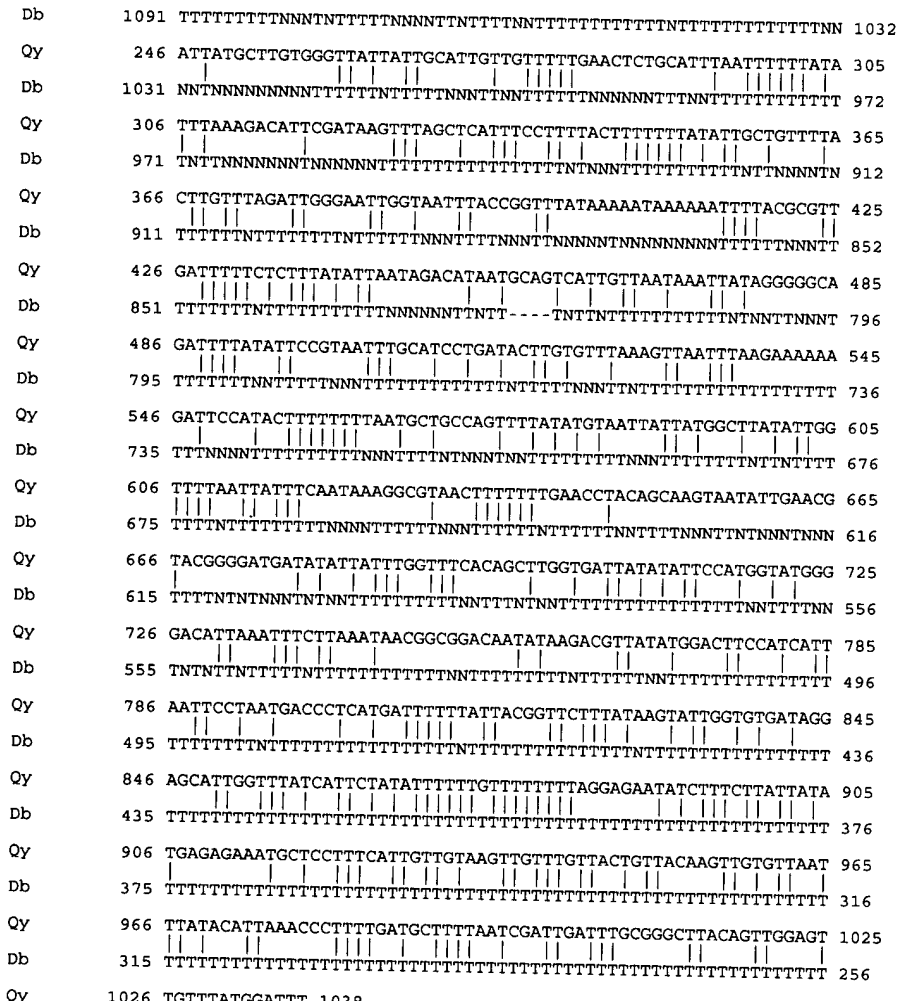
RESULT 3
DN705619/c
LOCUS DN705619 1299 bp mRNA linear EST 30-MAR-2005
DEFINITION CLJ60-F02.xld-t SHGC-CLJ Gasterosteus aculeatus cDNA clone CLJ60-F02 3', mRNA sequence.
ACCESSION DN705619
VERSION DN705619.1 GI:62070659
KEYWORDS EST.
SOURCE Gasterosteus aculeatus (three spined stickleback)
ORGANISM Gasterosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes; Gasterosteidae; Gasterosteus.
REFERENCE 1 (bases 1 to 1299)
AUTHORS Kingsley,D.M., Peichel,C., Balabahdra,S., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
TITLE Expressed sequence tags from Gasterosteus aculeatus
JOURNAL Unpublished (2003)
COMMENT Contact: Grimwood, Jane
Stanford Human Genome Center
Stanford University School of Medicine
975 S California Ave, Palo Alto, CA 94304, USA
Tel: 650 320 5917
Fax: 650 320 5801
Email: jane@shgc.stanford.edu
Plate: 60
High quality sequence start: 194
High quality sequence stop: 429.

FEATURES Location/Qualifiers
source 1. .1299
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Bitrufjordur marine sticklebacks, Iceland"
/db_xref="taxon:69293"
/clone="CLJ60-F02"
/sex="mixed male and female"
/tissue_type="whole larva"
/dev_stage="21 day old larvae collected at Swarup Stage 30 (J. Embryol. Exp. Morphol 6: 373-383.1958)"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="SHGC-CLJ"
/note="Vector: Express 1; Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by Open Biosystems. First strand cDNA synthesis was primed with an 54 bp linker primer containing an oligodT sequence preceeded by a synthetic NotI site (first strand primer: 5'-GACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3'). Following second strand synthesis, cDNAs were made blunt at the end corresponding to the original 5 prime end of mRNA, and cloned directionally into the NotI and EcoRV sites of Express 1. Note that the EcoRV site is typically destroyed in the blunt end cloning, leaving a junction of the form 'xxxATC'(where is ATC is the second half of the EcoRV site, and xxx is derived from the cDNA sequence). A map of the Express 1 vector is available at: http://www.openbiosystems.com/cdna_library_construction_faq.php#8 The primary library was transformed and amplified in DH10B (T1 phage resistant) bacteria. Clones available from Open Biosystems: http://www.openbiosystems.com/stickleback"

ORIGIN

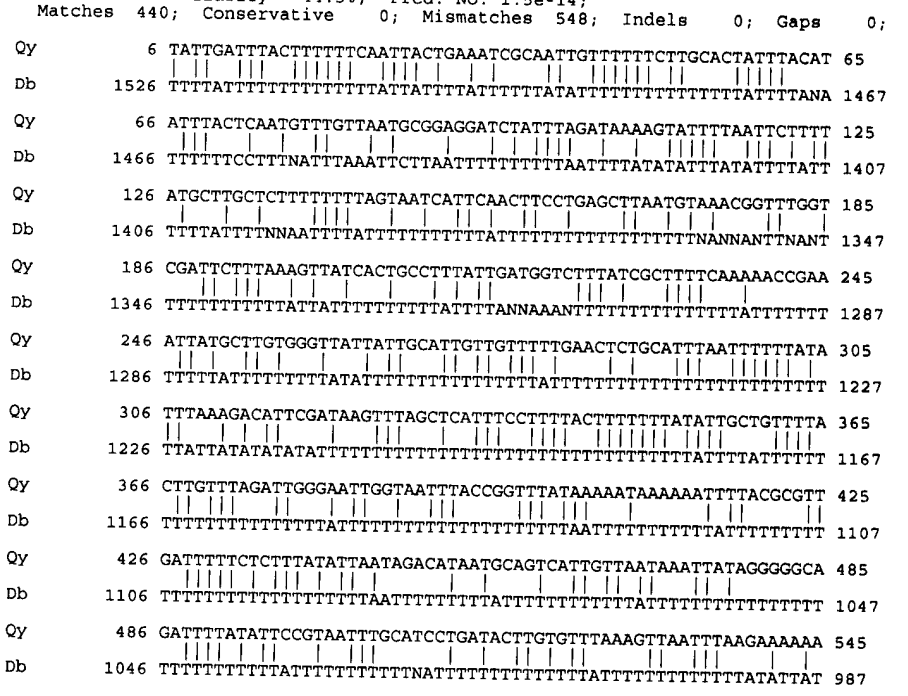
Query Match 13.2%; Score 138.8; DB 8; Length 1299;
Best Local Similarity 37.6%; Pred. No. 1.6e-14;
Matches 388; Conservative 0; Mismatches 641; Indels 4; Gaps 1;





RESULT 4
CF238805/c
LOCUS CF238805 1626 bp mRNA linear EST 05-AUG-2003
DEFINITION AGENCOURT_15099447 NICHD_XGC_Emb6 Xenopus tropicalis cDNA clone IMAGE:6995950 5', mRNA sequence.
ACCESSION CF238805
VERSION CF238805.1 GI:33442013
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1626)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Robert M. Grainger
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM14680 row: g column: 21
High quality sequence start: 71
High quality sequence stop: 316.
FEATURES Location/Qualifiers
source 1. .1626
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:6995950"
/tissue_type="neurula"
/dev_stage="embryo, stages 14-19"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD_XGC_Emb6"
/note="Vector: pCMV-SPORT6.1; Site_1: NotI; Site_2: EcoRV; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb. Constructed by Invitrogen. Note: This is a Xenopus Gene Collection (XGC) library."
ORIGIN

Query Match 13.2%; Score 138.8; DB 6; Length 1626;
Best Local Similarity 44.5%; Pred. No. 1.5e-14;
Matches 440; Conservative 0; Mismatches 548; Indels 0; Gaps 0;



```
Qy        546 GATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGG 605
              ||    |  ||||||||  |  |      ||| | | |  || || |   || | ||
Db        986 TTTTTTTTATTTTTTTTTTTTTTTTTTTATTTTTTTTTNTTTTTTTTTTTTTTTTTTT 927

Qy        606 TTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACG 665
              ||||  | || |   |       |   |||||||     ||      |    || |
Db        926 TTTTTTTAATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTAAATTTAATTTTTT 867

Qy        666 TACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATGGG 725
              |       | | || | | |  |||       ||  | | | | |||   |  | |
Db        866 TTNTNATTTATTTTTTTTTNTTTTTANTNNTTTTTTTTTAAAAAATTTTTTTTTTTTTT 807

Qy        726 GACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATT 785
                ||     | ||   |             | |      || | |  | || || |
Db        806 TTTTTTTTTNNTTTTNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAATTNNATAAAA 747

Qy        786 AATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATAGG 845
              ||  | |      | | ||||||  ||      | || || | ||||  | |  |
Db        746 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTACTTTTAATTATTATTTTTTTTTTTTT 687

Qy        846 AGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTATTATA 905
              |||  ||  |  || | | ||||||  |||||||       | | ||| || || | |
Db        686 NANATTTNTTNTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTT 627

Qy        906 TGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAAT 965
              |   | | ||  |  | | ||| || |   || ||| ||      |           ||
Db        626 TANNANANATAATTATATTATTTTTTTTTTTTTTTTTTTTTNNNNNANNNNAAAAAAAATAT 567

Qy        966 TTATACATTAAACCCTTTTGATGCTTTT 993
              ||||| |||      ||||  |  || |
Db        566 TTATAAATTTTTTTTTTTTTTTTTTNT 539
```

```
RESULT 5
CL078589/c
LOCUS       CL078589               1493 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-151K5_Sp5.1 CH216 Xenopus tropicalis genomic clone
            CH216-151K5, genomic survey sequence.
ACCESSION   CL078589
VERSION     CL078589.1  GI:40534502
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1493)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp5 atctgccgtttcgatcct
            Class: BAC ends
            High quality sequence start: 1009
            High quality sequence stop: 1070.
FEATURES             Location/Qualifiers
     source          1. .1493
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-151K5"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
```

```
ORIGIN

  Query Match           13.1%;  Score 137.8;  DB 10;  Length 1493;
  Best Local Similarity 46.1%;  Pred. No. 2.3e-14;
  Matches  465;  Conservative    0;  Mismatches  540;  Indels    3;  Gaps    2;

Qy          2 TGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTT 61
              | | ||||  | || |||||     ||  | || |      || || ||| ||       |||
Db       1312 TNTTTATTTTTATAATTTTTATTTTTTTTTAATTTTTTTTTTTTTATTTTTTTTTTATTTT 1253

Qy         62 ACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTC 121
                |||| |   | ||| |  ||       | | |||    | |   | ||||||||
Db       1252 TNTATTTTATTTTTTTTTTTTATATTTTATTTTTATTTTTTTTTTTATTTTTTTTAATTT 1193

Qy        122 TTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTT 181
              || |||  ||  | |||| |||  |  |  ||     ||  |     |||||  |      ||
Db       1192 TTATATATTTTTTTTTTTTATTTTATTTTTTTTTTTTTTATTTTTTTTAATTTTTTTAATT 1133

Qy        182 TGGTCGATTCTTTAA--AGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAA 239
              |  |  ||| |||    | || |    |  ||| ||  |  | ||| |   ||||  |
Db       1132 TTTTTTATTTTTTTTTTATTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTATTTTTTTTATT 1073

Qy        240 ACCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTTGAACTCTGCATTTAATTT 299
              |      ||||  || |   || || ||   || || |||||     |     |||  |||
Db       1072 ATTTTTTTTATTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTATTTTTTTTTT 1013

Qy        300 TTTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCT 359
              ||| | |||     |  ||    |   |||    | |||     | | |||||||  | ||  |
Db       1012 TTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTAATAATTTTTTTTTTTTTATT 953

Qy        360 GTTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTA 419
              | |  || |||  |||       ||  |  ||||    ||| |       |      ||||
Db        952 AATATTTTTTTTTTTATTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTT 893

Qy        420 CGCGTTGATTTTTCTC-TTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATA 478
                ||  | |||  |  ||||  ||  |      |  |     |  || ||  |  ||||
Db        892 TTTTTTTTTNTTTTTATTTTANTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATT 833

Qy        479 GGGGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAA 538
                  ||||  | ||    || |||    |  |   |  || | |||     || | ||||
Db        832 TTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTAATTAA 773

Qy        539 GAAAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTT 598
              | ||   ||    |  |||||||  |  |      |||| ||| || |||| |   ||
Db        772 ATATAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTATATATTTATTTTTTTTT 713

Qy        599 ATATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATA 658
              | ||  ||||  || |||   |        |   |||||||     |  |  ||  || |
Db        712 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTAATAAAAAAAAA 653

Qy        659 TTGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATG 718
                |   |       |  | | || ||   |||       |  |  || | | ||   |
Db        652 AAAANTTTTTTTNTTTTTTTTTTTTTTTNTTTTTTTTTTTTNTTNTTTTTTTTTTTTTTTT 593

Qy        719 GTATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTC 778
              | |       ||   ||| ||  |       |  |     | |  |    ||
Db        592 TTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTNTTTTTTTTTTTTTTTTTNTTTTTTTTTTTT 533

Qy        779 CATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTG 838
              | ||  ||  |  |   |  | ||||| ||    || ||  |  | | ||  |
Db        532 TTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTNNTTTTTNTTTTTT 473

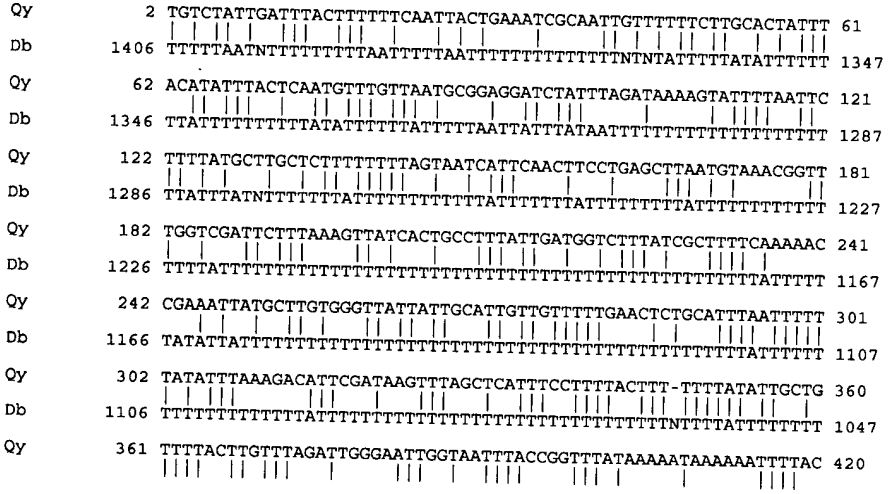
Qy        839 TGATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCT 898
              |  |       ||  ||| |  || |   |||||| ||||||||       | | ||| |
Db        472 TTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 413

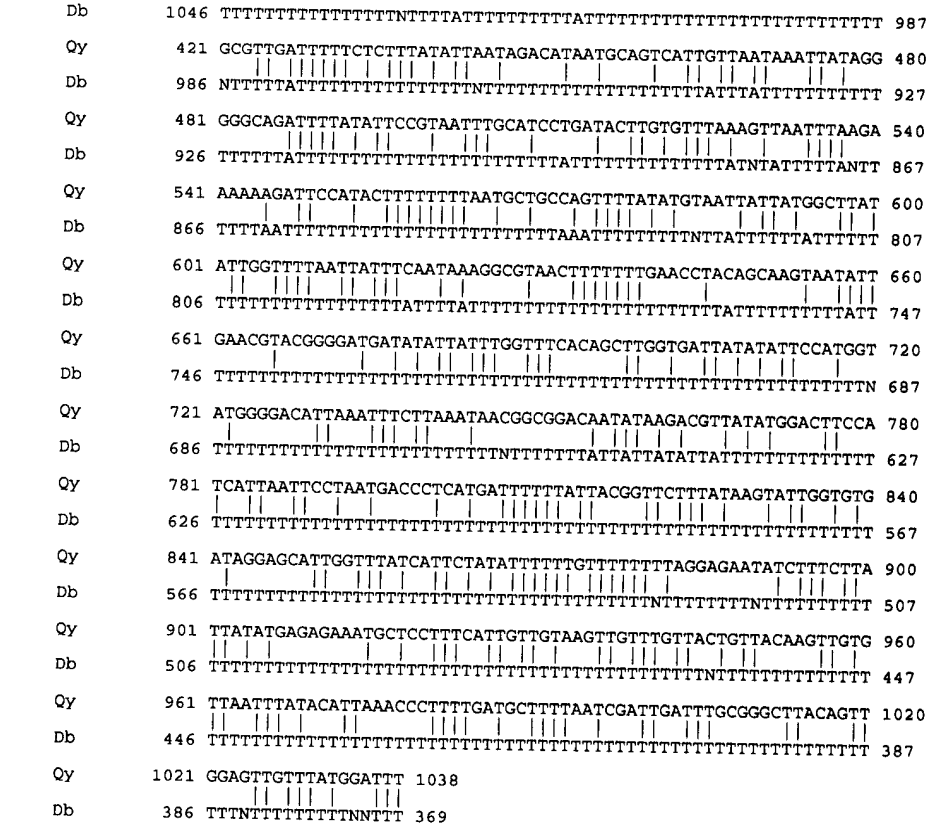
Qy        899 TATTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTG 958
              | || | |          | |  |||  || || |   || ||| ||  | ||      ||
Db        412 TTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTNNTTTTT 353

Qy        959 TGTTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTT 1006
              | ||  ||| |   ||      ||||  |  |||       ||  |||
Db        352 TTTTTTTTTTTTTTTTTNTTTTTTTTNNTTTTTTNNTNNTTTTNTTTT 305
```

RESULT 6
CL078538/c
LOCUS CL078538 1536 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-151C11_Sp5.1 CH216 Xenopus tropicalis genomic clone CH216-151C11, genomic survey sequence.
ACCESSION CL078538
VERSION CL078538.1 GI:40534451
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1536)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp5 atctgccgtttcgatcct
Class: BAC ends
High quality sequence start: 1061
High quality sequence stop: 1145.
FEATURES Location/Qualifiers
source 1. .1536
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-151C11"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"
ORIGIN

Query Match 13.1%; Score 137.6; DB 10; Length 1536;
Best Local Similarity 45.8%; Pred. No. 2.5e-14;
Matches 475; Conservative 0; Mismatches 562; Indels 1; Gaps 1;





RESULT 7
AJ592058
LOCUS AJ592058 1434 bp DNA linear GSS 15-JAN-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, right border, clone 602D05, genomic survey sequence.
ACCESSION AJ592058
VERSION AJ592058.1 GI:37941682
KEYWORDS GSS; right border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F., Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G., Lepiniec,L., Caboche,M. and Lecharny,A.
TITLE T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)
PUBMED 12446565
REFERENCE 2 (bases 1 to 1434)
AUTHORS Balzergue,S.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE

```
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .1434
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="602D05"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Wassilewskija"
     misc_feature    1. .1434
                     /note="T-DNA flanking sequence
                     right border"
ORIGIN

  Query Match             13.0%;  Score 136.8;  DB 10;  Length 1434;
  Best Local Similarity   43.6%;  Pred. No. 3.5e-14;
  Matches  450;  Conservative    0;  Mismatches  583;  Indels    0;  Gaps    0;

Qy          6 TATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTACAT 65
              | |   |||  ||||||   ||  |     |     ||  ||||| ||    | |||   |
Db        157 TTTNTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTNTTTTTTTTTNTNTTTTTTTT 216

Qy         66 ATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTT 125
                |||  |   | ||| ||  |       | | |||   |      | ||||  || ||||
Db        217 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTTTTTTTTTTTTTTTTTTTTT 276

Qy        126 ATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGT 185
                |  ||  | ||||||||  |  |  ||    ||  |    ||  | |      |||  |
Db        277 TTTTTTTTTTTTTTTTTTTTTTTNTTTTNNTNTTTTTTTTTTTTTTTTTTTTTTTTNTT 336

Qy        186 CGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAA 245
                || |||    || |   |   ||| ||  |  | ||| |   ||||
Db        337 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 396

Qy        246 ATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATA 305
                || |  || |   || || ||   || || |||||    | |   |||  |||||| |
Db        397 TTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 456

Qy        306 TTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTA 365
                |||      ||   |  |||   |  |||  ||||  ||||||| | ||  | ||||
Db        457 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTT 516

Qy        366 CTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTT 425
                || |||   ||     ||  |   ||     ||| |      |      ||||     ||
Db        517 TTTTTTTTTTTTTTTTTTTTTTTNTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 576

Qy        426 GATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCA 485
                |||||  | |||  | ||  |      |  |      |  || ||  |   || |
Db        577 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTT 636

Qy        486 GATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAA 545
                ||||  | ||   |  |||   |  |  |  || | |||    ||  |||
Db        637 TTTTTTTTTTTTTTTNTTTTTTTTNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 696

Qy        546 GATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGG 605
                ||   |  ||||||||  |   |     |||| | | |  || || |   || | ||
Db        697 TTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 756

Qy        606 TTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACG 665
                ||||  || |||   |       |   |||||||     |         |  | ||
Db        757 TTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTT 816

Qy        666 TACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATGGG 725
```

```
                |       |  | | || |||  |||      ||  |  || | | ||   |  | |
Db        817 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTT 876

Qy        726 GACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATT 785
                ||   ||| ||   |                | |      | | |     ||    |  ||
Db        877 TTTTTTTTTTTTTTTTTNTTTTTTTTTNTTTTTTTTNTTTNNTTTTTTTTTTTTTTTTT 936

Qy        786 AATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATAGG 845
                ||  |  |      |  |  ||||||  |    |    | |  | ||  | |  |
Db        937 TTTNTTTTTTTNTTTTTTTTTTTTTTTNTTTTTNTNNTTTTTTTTTTNNTTTTTTTNT 996

Qy        846 AGCATTGGTTTATCATTCTATATTTTTGTTTTTTTTAGGAGAATATCTTTCTTATTATA 905
                ||  ||| |  || | | ||||||  |||||||         | | ||| || || |
Db        997 NTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTNNTTTTTTTTTTTTTTTT 1056

Qy        906 TGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAAT 965
              |         |  |  |||  || || |   || ||| ||  | ||     || | ||  |
Db       1057 TTTTTTTTTNTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTT 1116

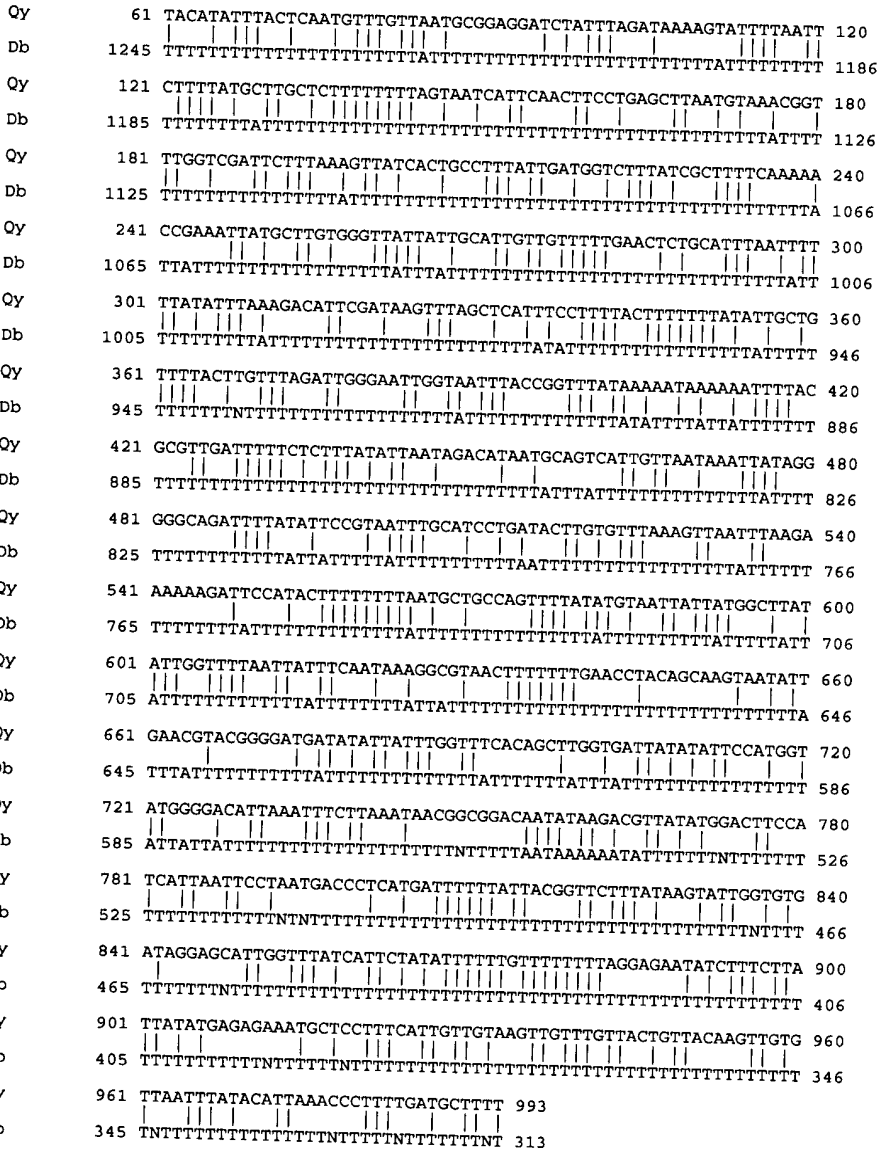
Qy        966 TTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTTGGAGT 1025
              || |   ||      || |  |  |||   |   |  |||     ||     |     |
Db       1117 TTTTTTTTTNNTTTTTNTNNTTTTTNTTTTTNTNNTTTTTTTTNNTTTTTTNNTTTT 1176

Qy       1026 TGTTTATGGATTT 1038
              | ||| |   | |
Db       1177 TNTTTTTTTNTNT 1189
```

```
RESULT 8
CL083840/c
LOCUS       CL083840                1599 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-2H14_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-2H14,
            genomic survey sequence.
ACCESSION   CL083840
VERSION     CL083840.1  GI:40558745
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1599)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 1069
            High quality sequence stop: 1158.
FEATURES             Location/Qualifiers
     source          1. .1599
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-2H14"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             13.0%;  Score 136.8;  DB 10;  Length 1599;
  Best Local Similarity   45.6%;  Pred. No. 3.5e-14;
  Matches  453;  Conservative    0;  Mismatches  540;  Indels    0;  Gaps    0;

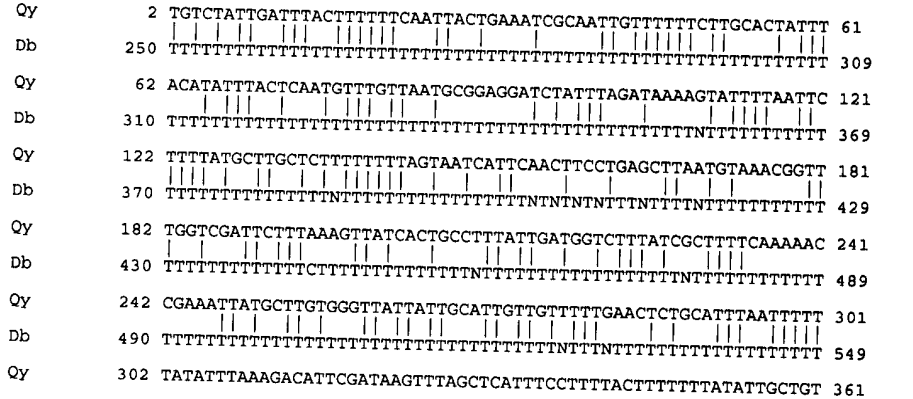
Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              || | | ||  |||  ||||||   ||  |     |     ||  | |||| ||    | ||
Db       1305 ATTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTATTTTTTTATTTTTTTTTTTTT 1246
```

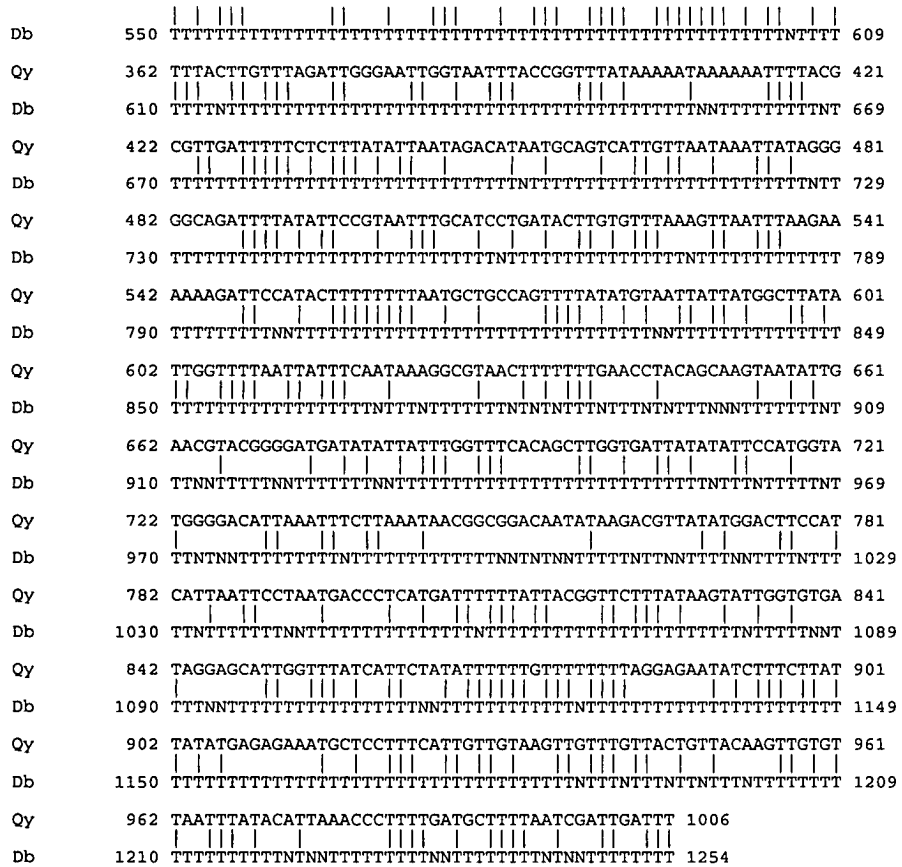


RESULT 9
AG346189
LOCUS AG346189 1507 bp DNA linear GSS 18-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-141E08.TJ, genomic survey sequence.
ACCESSION AG346189
VERSION AG346189.1 GI:47919499

KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1507)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1507
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-141E08.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

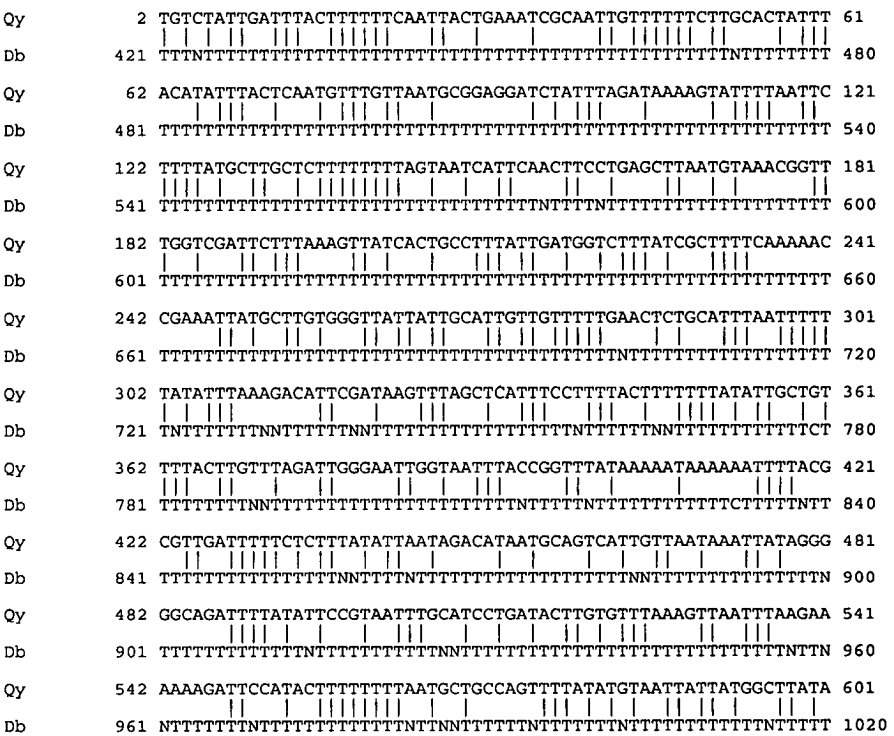
Query Match 13.0%; Score 136.6; DB 10; Length 1507;
Best Local Similarity 42.8%; Pred. No. 3.8e-14;
Matches 430; Conservative 0; Mismatches 575; Indels 0; Gaps 0;





RESULT 10
AG386893
LOCUS AG386893 1519 bp DNA linear GSS 21-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-201E07.TJ, genomic survey sequence.
ACCESSION AG386893
VERSION AG386893.1 GI:47998098
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1519)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1519
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-201E07.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

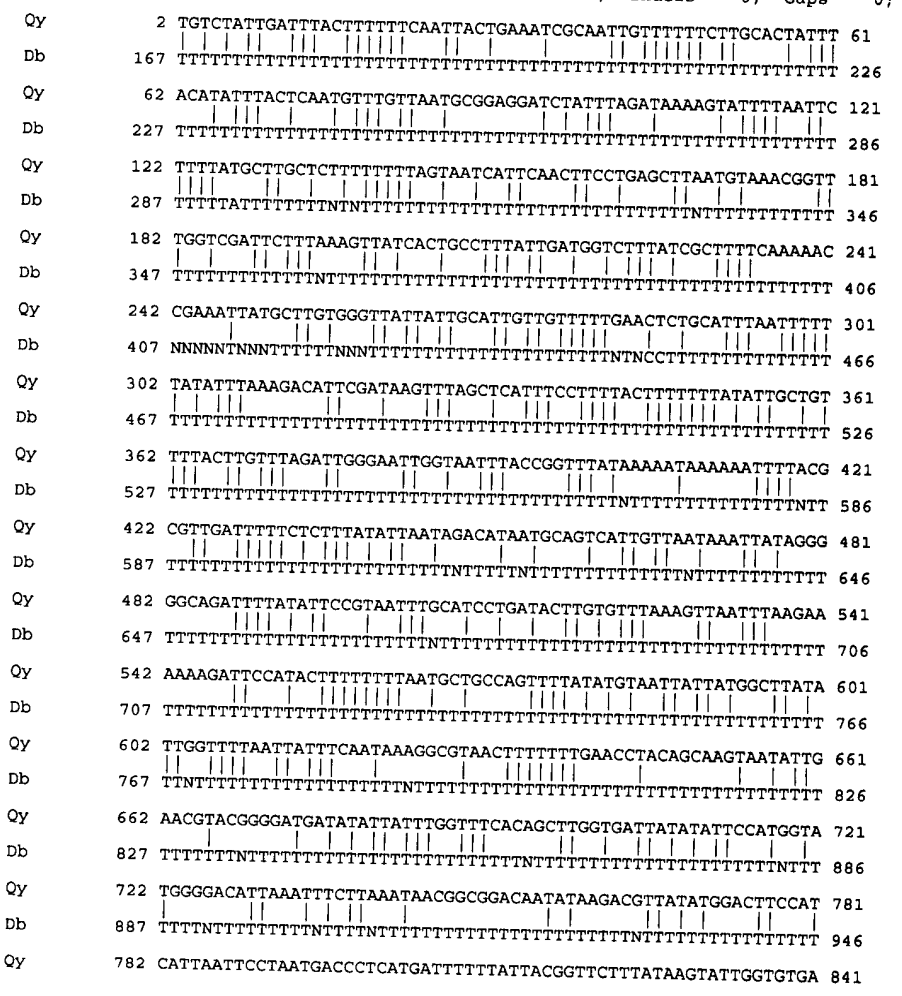
Query Match 12.9%; Score 136.2; DB 10; Length 1519;
Best Local Similarity 42.2%; Pred. No. 4.4e-14;
Matches 438; Conservative 0; Mismatches 599; Indels 0; Gaps 0;

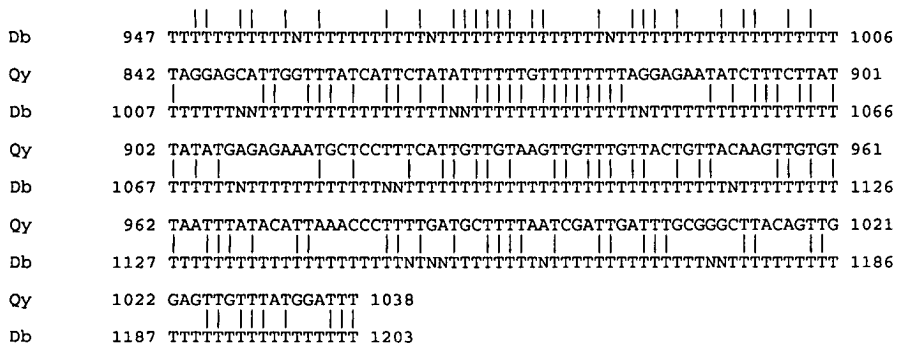




RESULT 11
AG341525
LOCUS AG341525 1334 bp DNA linear GSS 18-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-134O22.TJ, genomic survey sequence.
ACCESSION AG341525
VERSION AG341525.1 GI:47914835
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1334)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp). Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1334
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-134O22.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Query Match 12.9%; Score 135.8; DB 10; Length 1334;
Best Local Similarity 43.9%; Pred. No. 5.3e-14;
Matches 455; Conservative 0; Mismatches 582; Indels 0; Gaps 0;





RESULT 12
CL038406/c
LOCUS CL038406 1594 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-46A6_Sp6.1 CH216 Xenopus tropicalis genomic clone CH216-46A6, genomic survey sequence.
ACCESSION CL038406
VERSION CL038406.1 GI:40494319
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1594)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 471
High quality sequence stop: 522.
FEATURES Location/Qualifiers
source 1. .1594
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-46A6"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"
ORIGIN

Query Match 12.9%; Score 135.8; DB 10; Length 1594;
Best Local Similarity 45.6%; Pred. No. 5.2e-14;
Matches 458; Conservative 0; Mismatches 547; Indels 0; Gaps 0;

RESULT 13
CL082650/c
LOCUS CL082650 1367 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-168O1_Sp5.1 CH216 Xenopus tropicalis genomic clone CH216-168O1, genomic survey sequence.
ACCESSION CL082650
VERSION CL082650.1 GI:40538563
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;

```
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1367)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp5 atctgccgtttcgatcct
            Class: BAC ends
            High quality sequence start: 292
            High quality sequence stop: 1179.
FEATURES             Location/Qualifiers
     source          1. .1367
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-168O1"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             12.8%;  Score 135.2;  DB 10;  Length 1367;
  Best Local Similarity   45.6%;  Pred. No. 6.7e-14;
  Matches  473;  Conservative    0;  Mismatches  564;  Indels    0;  Gaps    0;

Qy          2 TGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTT 61
Db       1298 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTT 1239

Qy         62 ACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTC 121
Db       1238 TTTTTTTTTTTTTTTTTATATTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTATTT 1179

Qy        122 TTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTT 181
Db       1178 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTT 1119

Qy        182 TGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAAC 241
Db       1118 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1059

Qy        242 CGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTT 301
Db       1058 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTT 999

Qy        302 TATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGT 361
Db        998 TTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 939

Qy        362 TTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACG 421
Db        938 TTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTATTTTTTTTTTTTTT 879

Qy        422 CGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGG 481
Db        878 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 819

Qy        482 GGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAA 541
Db        818 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 759

Qy        542 AAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATA 601
Db        758 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 699

Qy        602 TTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTG 661
Db        698 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 639

Qy        662 AACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTA 721
Db        638 TTTTTATTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 579

Qy        722 TGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCAT 781
Db        578 TTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 519

Qy        782 CATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGA 841
Db        518 TTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 459

Qy        842 TAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTAT 901
Db        458 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 399

Qy        902 TATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGT 961
Db        398 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 339

Qy        962 TAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTTG 1021
Db        338 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 279

Qy       1022 GAGTTGTTTATGGATTT 1038
Db        278 TTTTTTTTTTTTTTTTT 262


RESULT 14
DN698671/c
LOCUS       DN698671                1447 bp    mRNA    linear   EST 30-MAR-2005
DEFINITION  CLJ18-H11.x1d-t SHGC-CLJ Gasterosteus aculeatus cDNA clone
            CLJ18-H11 3', mRNA sequence.
ACCESSION   DN698671
VERSION     DN698671.1  GI:62060149
KEYWORDS    EST.
SOURCE      Gasterosteus aculeatus (three spined stickleback)
  ORGANISM  Gasterosteus aculeatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
            Gasterosteidae; Gasterosteus.
REFERENCE   1  (bases 1 to 1447)
  AUTHORS   Kingsley,D.M., Peichel,C., Balabahdra,S., Grimwood,J., Dickson,M.,
            Schmutz,J. and Myers,R.M.
  TITLE     Expressed sequence tags from Gasterosteus aculeatus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Grimwood, Jane
            Stanford Human Genome Center
            Stanford University School of Medicine
            975 S California Ave, Palo Alto, CA 94304, USA
            Tel: 650 320 5917
            Fax: 650 320 5801
            Email: jane@shgc.stanford.edu
            Plate: 18
            High quality sequence start: 93
            High quality sequence stop: 294.
FEATURES             Location/Qualifiers
     source          1. .1447
                     /organism="Gasterosteus aculeatus"
                     /mol_type="mRNA"
                     /strain="Bitrufjordur marine sticklebacks, Iceland"
                     /db_xref="taxon:69293"
                     /clone="CLJ18-H11"
                     /sex="mixed male and female"
                     /tissue_type="whole larva"
                     /dev_stage="21 day old larvae collected at Swarup Stage 30
                     (J. Embryol. Exp. Morphol 6: 373-383.1958)"
```

/lab_host="DH10B (T1 phage resistant)"
/clone_lib="SHGC-CLJ"
/note="Vector: Express 1; Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by Open Biosystems. First strand cDNA synthesis was primed with an 54 bp linker primer containing an oligodT sequence preceeded by a synthetic NotI site (first strand primer: 5'-GACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3'). Following second strand synthesis, cDNAs were made blunt at the end corresponding to the original 5 prime end of mRNA, and cloned directionally into the NotI and EcoRV sites of Express 1. Note that the EcoRV site is typically destroyed in the blunt end cloning, leaving a junction of the form 'xxxATC' (where is ATC is the second half of the EcoRV site, and xxx is derived from the cDNA sequence). A map of the Express 1 vector is available at: http://www.openbiosystems.com/cdna_library_construction_faq.php#8 The primary library was transformed and amplified in DH10B (T1 phage resistant) bacteria. Clones available from Open Biosystems: http://www.openbiosystems.com/stickleback"

ORIGIN

Query Match 12.8%; Score 134.8; DB 8; Length 1447;
Best Local Similarity 37.7%; Pred. No. 7.8e-14;
Matches 391; Conservative 0; Mismatches 646; Indels 0; Gaps 0;

```
Qy          2 TGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTT 61
Db       1172 TGTGGNTTNNTTTNTNTTTTTTTTTTNNTTNTNNTTTNTTTTTTTTTTTTTTTTTNTTT 1113

Qy         62 ACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTC 121
Db       1112 TTTTTNNNNTTTTTTNNNTNNNNTTTTTTNNNNNNTTTTTTTTTNNTTTTTNNNNNTTTT 1053

Qy        122 TTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTT 181
Db       1052 TTTTTTTTTTNTTTTTTTTNNNTNTTTTTNNNNNNTNNNTTTTTTTTTTTTTTTTTTTT 993

Qy        182 TGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAAC 241
Db        992 TTTTNNTTNNNTTNNNNNNNNNTTNNNTTTTTTTTTTTTTTTTTTTTTTNNNNTTTTTTN 933

Qy        242 CGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTT 301
Db        932 NNTNTTTTTTTTTNTTTTTTTTTTTTTTTNNNNTTTNTTTTTNNNNTTTTTTTTTTGGTT 873

Qy        302 TATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGT 361
Db        872 TNTNTTNNNNNNNNNTTTTTNTNTTTTTTTNNTTTTTTTTTTTTTTTTTTTTNNNTTNTTTT 813

Qy        362 TTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACG 421
Db        812 NTNNNTTTNTTTNNNNTTTTNNTTTTNTTTTTTTTNNTTTTNNNNTTTTTTTNTTTTTNTN 753

Qy        422 CGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGG 481
Db        752 TTNTTNNTTTTTNTNNTTTNTTTTTTTTTTTTTNTTTNTTTTNNTNTTNNTNTTTTTTT 693

Qy        482 GGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAA 541
Db        692 TTTTTTTNTTTNTTTTTTTTNNTTTNTTTNTNTTTTTTTNNNTTTNTNTTTTTNTTNTTT 633

Qy        542 AAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATA 601
Db        632 TTNNTTTTNTTTTTTTTNTTTTTTTTNTTTNNTTTTTTNTTTTTTTTTNNTTTTTNTTT 573

Qy        602 TTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTG 661
Db        572 TTTTTTTTTTTTTTTTNTNTTTGGGTTNTTTTTTTTTTTNNNTTTTTTTTTTNTTNTNTTN 513

Qy        662 AACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTA 721
```

```
Db        512 TNTNTTNTTGGGTTNNNTTTTNTTTTTTTNNNNNTTTTTTTTTNTTTTTNTTNNTTTNTNT 453

Qy        722 TGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCAT 781
Db        452 TTTTNNTNTTNTTTTTTTNNNNNTNNTNTTTTTTTTTTTTTTTNTTTTTTNTNNNTTTTTT 393

Qy        782 CATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGA 841
Db        392 NTNTTTTTNTTTTTTTTTTTTTTNTTTNNTTNTTNTTTTTTTTTTTTTTTTTTTTTTTT 333

Qy        842 TAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTAT 901
Db        332 TTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 273

Qy        902 TATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGT 961
Db        272 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 213

Qy        962 TAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTTG 1021
Db        212 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 153

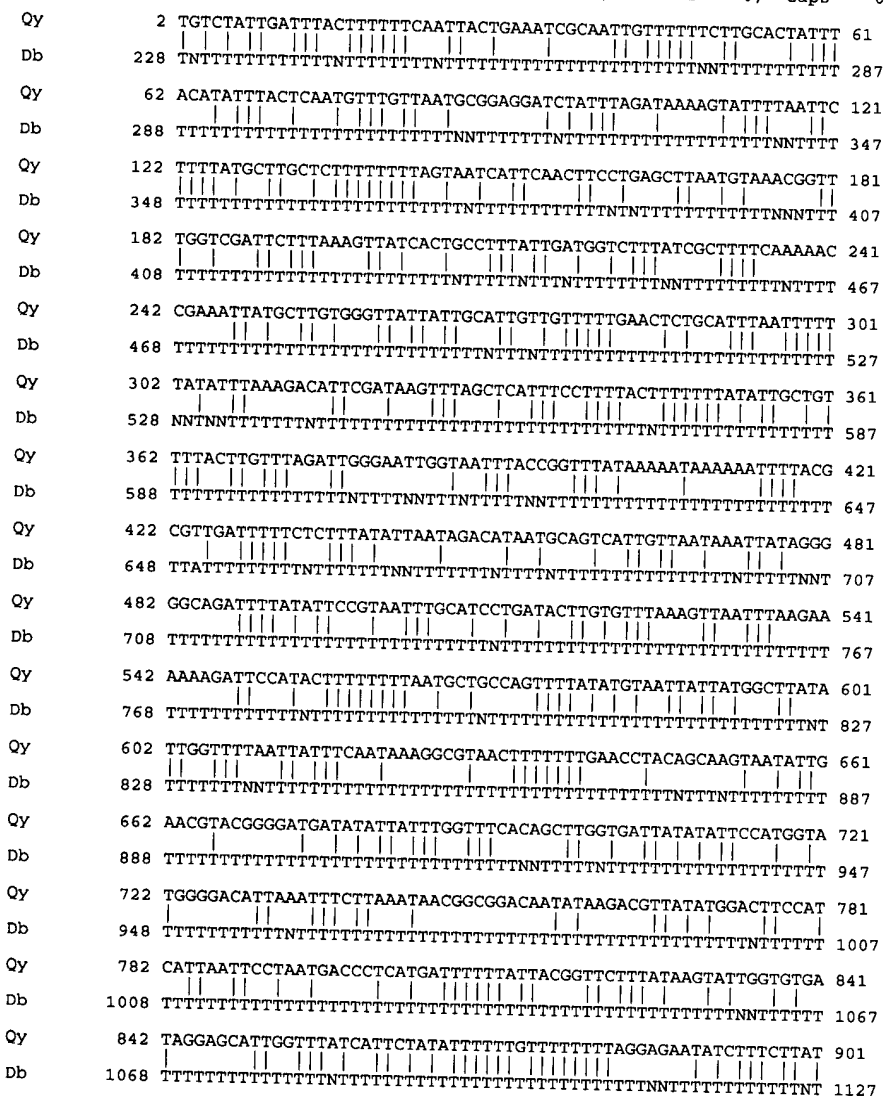
Qy       1022 GAGTTGTTTATGGATTT 1038
Db        152 TTTTTTTTTTTTTTTTT 136
```

RESULT 15
AG332292
LOCUS AG332292 1482 bp DNA linear GSS 18-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-122G19.TJ, genomic survey sequence.
ACCESSION AG332292
VERSION AG332292.1 GI:47905602
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
PUBMED 15574823
REFERENCE 2 (bases 1 to 1482)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .1482

/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-122G19.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

Query Match 12.8%; Score 134.8; DB 10; Length 1482;
Best Local Similarity 42.9%; Pred. No. 7.8e-14;
Matches 445; Conservative 0; Mismatches 592; Indels 0; Gaps 0;





Search completed: May 11, 2006, 05:00:53
Job time : 3375.55 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 20:34:42 ; Search time 146.879 Seconds
(without alignments)
12743.667 Million cell updates/sec

Title: US-09-423-093B-1_COPY_9901_10953
Perfect score: 1053
Sequence: 1 atgtctattgatttactttt.....gatttgcaaaaattagataa 1053

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	95.6	9.1	1039	3	US-09-902-540-1280	Sequence 1280, Ap
c	2	95	9.0	612	3	US-09-902-540-1357	Sequence 1357, Ap
c	3	79.6	7.6	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	4	75.6	7.2	700	3	US-09-735-271-1038	Sequence 1038, Ap
c	5	73	6.9	187169	3	US-09-949-016-12776	Sequence 12776, A
c	6	73	6.9	191569	3	US-09-949-016-15940	Sequence 15940, A
	7	69.8	6.6	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	8	67.8	6.4	1132	3	US-09-533-559-5452	Sequence 5452, Ap
c	9	65.8	6.2	614	3	US-09-902-540-1318	Sequence 1318, Ap
	10	64.6	6.1	451924	3	US-09-949-016-12896	Sequence 12896, A
	11	64.6	6.1	451925	3	US-09-949-016-17305	Sequence 17305, A
	12	63.4	6.0	50000	3	US-09-662-254B-24	Sequence 24, Appl
	13	63.4	6.0	147382	3	US-09-949-016-14624	Sequence 14624, A
	14	63	6.0	50000	3	US-09-662-254B-23	Sequence 23, Appl
c	15	62.2	5.9	205044	3	US-09-949-016-15851	Sequence 15851, A
c	16	62.2	5.9	205044	3	US-09-949-016-15852	Sequence 15852, A
c	17	62.2	5.9	205044	3	US-09-949-016-15853	Sequence 15853, A
c	18	62.2	5.9	223471	3	US-09-949-016-12387	Sequence 12387, A
c	19	62.2	5.9	223471	3	US-09-949-016-12724	Sequence 12724, A
c	20	62.2	5.9	223471	3	US-09-949-016-12725	Sequence 12725, A
	21	62	5.9	837	3	US-08-998-416-288	Sequence 288, App
	22	61.2	5.8	658	3	US-08-998-416-595	Sequence 595, App
	23	61	5.8	50000	3	US-09-662-254B-25	Sequence 25, Appl
	24	60.4	5.7	1055	3	US-09-806-708B-23	Sequence 23, Appl
	25	60.4	5.7	29717	3	US-09-949-016-16284	Sequence 16284, A
	26	60.4	5.7	60376	3	US-09-949-016-12423	Sequence 12423, A
	27	60	5.7	615	3	US-08-998-416-186	Sequence 186, App
c	28	59.4	5.6	10640	3	US-09-417-485D-5	Sequence 5, Appli
c	29	59.4	5.6	18773	3	US-09-949-016-14164	Sequence 14164, A
	30	58.6	5.6	7218	2	US-08-232-463-14	Sequence 14, Appl
c	31	58.6	5.6	119153	3	US-09-949-016-12378	Sequence 12378, A
c	32	58.4	5.5	95255	3	US-09-949-016-17067	Sequence 17067, A
	33	58.2	5.5	134987	3	US-09-949-016-15348	Sequence 15348, A
	34	58.2	5.5	134987	3	US-09-949-016-15349	Sequence 15349, A
	35	58.2	5.5	134987	3	US-09-949-016-15350	Sequence 15350, A
	36	58.2	5.5	134987	3	US-09-949-016-15507	Sequence 15507, A
	37	58.2	5.5	134987	3	US-09-949-016-15508	Sequence 15508, A
	38	58.2	5.5	134987	3	US-09-949-016-15509	Sequence 15509, A
	39	58	5.5	636	3	US-08-998-416-1137	Sequence 1137, Ap
c	40	57.8	5.5	387902	3	US-09-949-016-14543	Sequence 14543, A
c	41	57.8	5.5	421883	3	US-09-949-016-12557	Sequence 12557, A
	42	57.6	5.5	32392	3	US-09-662-254B-27	Sequence 27, Appl
	43	57.6	5.5	231129	3	US-09-949-016-16110	Sequence 16110, A
	44	57.6	5.5	266293	3	US-09-949-016-11934	Sequence 11934, A
	45	57	5.4	601	3	US-09-949-016-196365	Sequence 196365,

ALIGNMENTS

RESULT 1
US-09-902-540-1280/c
; Sequence 1280, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1280
; LENGTH: 1039
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1039)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1280

Query Match 9.1%; Score 95.6; DB 3; Length 1039;
Best Local Similarity 44.8%; Pred. No. 2.6e-10;
Matches 452; Conservative 0; Mismatches 546; Indels 10; Gaps 2;

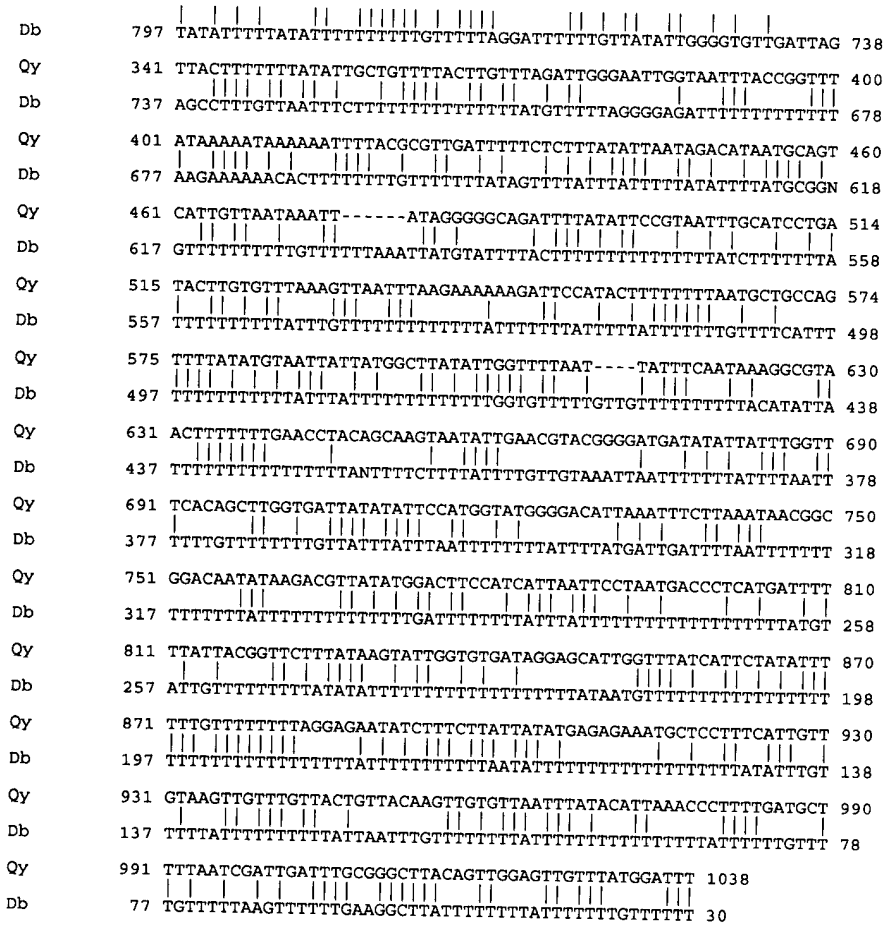
```
Qy        41 TTGTTTTTTCTTGCACTATTTACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATT 100
             || |||||| ||     | |||   | |||  |  |  ||| ||  |        ||   ||
Db      1037 TTTTTTTTTTTTATTTTGTTTTTTTTTTTTTTGTAATTTTTTTTTTTTTTTTTTATTAGTT 978

Qy       101 TAGATAAAAGTATTTTAATTCTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTC 160
             |          | ||||  || || |     ||  | |||||||||  |    |||
Db       977 TTATGTTTTTTTTTTTTTTTTTTTATGGTTTTATTTTTTTTTTTTTTGGATATTTTGTAAT 918

Qy       161 CTGAGCTTAATGTAAACGGTTTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGG 220
             |    ||    | |  |||  |   || |||    || |   |   || | | ||
Db       917 TTTTTTTTTTCTTTATTTTTTTATTTTTTATTTGGTTTTTTATTTTTTTTAAGTTATTT 858

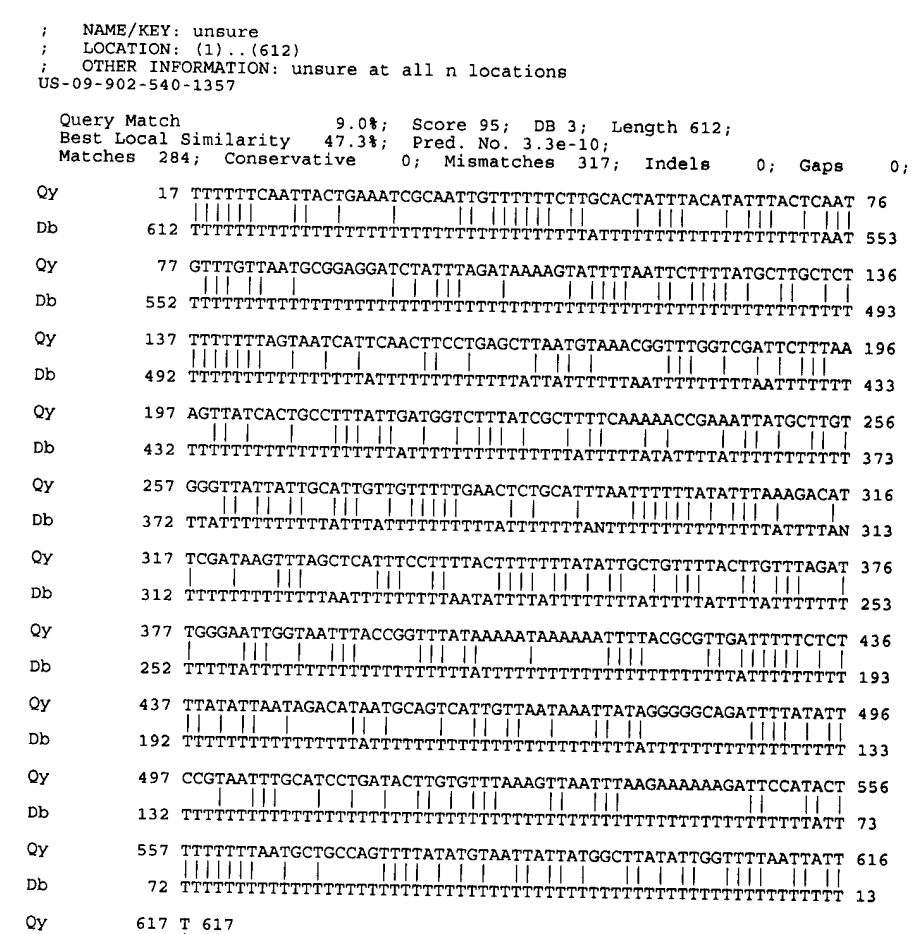
Qy       221 TCTTTATCGCTTTTCAAAAACCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTT 280
             |  || |  |||||    |       |  || || |  || || ||  ||       ||
Db       857 TTGTTTTTTATTTTCCCTCATTTTTTTAGTGTTTTTAATTTTTTTTTTTTTTTAGAAGGTT 798

Qy       281 TGAACTCTGCATTTAATTTTTTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTT 340
```



RESULT 2
US-09-902-540-1357/c
; Sequence 1357, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1357
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(612)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1357

Query Match 9.0%; Score 95; DB 3; Length 612;
Best Local Similarity 47.3%; Pred. No. 3.3e-10;
Matches 284; Conservative 0; Mismatches 317; Indels 0; Gaps 0;



RESULT 3
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellems, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California

```
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13

  Query Match             7.6%;  Score 79.6;  DB 2;  Length 19124;
  Best Local Similarity   47.5%;  Pred. No. 5.9e-07;
  Matches  304;  Conservative    0;  Mismatches  329;  Indels    7;  Gaps    2;

Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              ||| ||| || ||  | | |   |   || ||  |   |   |  || | |   |||
Db      16059 ATGGGTATATATATATATATATATGTATGTGTAAATGTTGTATGTAGTTATGGGTATATA 16000

Qy         61 TACATATTTA-CTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAAT 119
              || |||| ||  |   | | |||  |||      || |  ||  ||     |  ||||||
Db      15999 TATATATATATGTATGTATATGTGCATGTAATTAATTTTATTTTATTTTTTTTATTTAAT 15940

Qy        120 TCTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGG 179
              |   || |  |   | |||||| ||        |||    |   |    || || |||
Db      15939 TAAATTTTTTTATTTTTTTTTTATTTCATTAAATTTTTTTAATTTTTTTTTATTTAATTAA 15880

Qy        180 TTTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAA 239
              ||  |  |||  |||   || |   |   |||    |  |  ||    ||||
Db      15879 ATTTTTTTATTTATTATTTTTTTTTTATTTAATTAAATTTTTTTATTTATTATTTTTTTTT 15820

Qy        240 ACCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTT 299
              |  |||| |   || |   || |   |   ||  || ||||| |||  |   || | |||
Db      15819 ATTTAAATAAATTTTTTTTATTTTATGTATATATTTTTTTTTTAACATTTTTTTAATTTT 15760

Qy        300 TTTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCT 359
              ||| ||||| | || ||    |    ||     || |  ||||  |||||||       |
Db      15759 TTTTTATTTTATGATATATATTTTTATTTTAATATATTTTTTTCTTTTTTTT------TT 15706

Qy        360 GTTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTA 419
              |||||  |  | || |||     ||  | | |     ||| |     |     |||
Db      15705 GTTTTTATGATATATATTTTTTTTTTTTTTAATGTTTTTTTTTTTTCTTCTTTTGTTTTTAT 15646

Qy        420 CGCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAG 479
              || ||  |  | ||| | |||     | ||   |     | ||| ||  |   || ||
Db      15645 TTTTTTTATAATCATTTTTTTTTTTATATAAAATTTTTTTTTAATTTTTTTTTGATAATCT 15586

Qy        480 GGGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAG 539
              ||  ||||||  |    || ||      | ||  |  | ||||   ||  || ||
Db      15585 TTTTCATTTTTTATTCTATCAAAATTTATATTTTATTATAATTTTTATTATTTTTTAAAA 15526

Qy        540 AAAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTA 599
              ||      | |  | ||||||||  |  |      ||    |  |  || ||||   | |
Db      15525 AATTTTTCTCCTTTTTTTTTTTTTTTTTTTTTTATTTTAATAAATTTTTTTTTATATTTCA 15466

Qy        600 TATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTT 639
              | ||  ||||  | |||| ||||||   ||   | | |||
Db      15465 TTTTTCTTTTTTTCATTTTAATAAATTTGTTTTTATATTT 15426


RESULT 4
US-09-735-271-1038/c
; Sequence 1038, Application US/09735271
; Patent No. 6869762
; GENERAL INFORMATION:
;  APPLICANT: Daly, Mark J.
;  APPLICANT:  Hudson, Thomas J.
;  APPLICANT:  Lander, Eric S.
;  APPLICANT:  Rioux, John
;  APPLICANT:  Siminovitch, Kathy
;  TITLE OF INVENTION: IBD-RELATED POLYMORPHISMS
;  FILE REFERENCE: 2825.1025-002
;  CURRENT APPLICATION NUMBER: US/09/735,271
;  CURRENT FILING DATE:  2000-12-11
;  PRIOR APPLICATION NUMBER: US 60/170,257
;  PRIOR FILING DATE: 1999-12-10
;  PRIOR APPLICATION NUMBER: US 60/196,046
;  PRIOR FILING DATE: 2000-04-10
;  NUMBER OF SEQ ID NOS: 2058
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1038
;   LENGTH: 700
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(700)
;   OTHER INFORMATION: n = A,T,C or G
US-09-735-271-1038

  Query Match             7.2%;  Score 75.6;  DB 3;  Length 700;
  Best Local Similarity   46.3%;  Pred. No. 2.5e-06;
  Matches  283;  Conservative    0;  Mismatches  325;  Indels    3;  Gaps    1;

Qy        296 ATTTTTTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATAT 355
              |||   ||||| | || | |||  |||  | |   |     |  ||  |||  ||||
Db        688 ATTAAATATATATTAATATATTAAATATATTAATATAATTAATATATATTTAATATAA 629

Qy        356 TGCTGTTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAAT 415
              |    | | |    | || |||    |||   | |       | ||||   | || | ||
Db        628 TNTAATATAATAAATATATATTATATATTAAATAATATATATTATATATTTAAAATATAT 569

Qy        416 TTTACGCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATT 475
              |||      |  | | | | ||||    ||||  | |  ||  | | |   |  ||||  |
Db        568 ATTAATATATATTATATATATTTA---AAATATATAATATATATTAAACATATTTAAAAT 512

Qy        476 ATAGGGGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATT 535
              |||       | | | || |||   || | |  ||  |     || | | | | |  | |
Db        511 ATATATTTAAAAATATAAATTATATATTATATATAATATATATTATATATTATATAAT 452

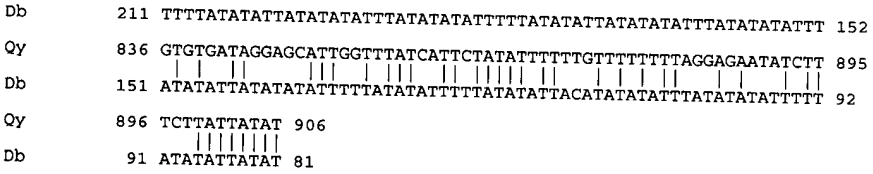
Qy        536 TAAGAAAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGG 595
              | || | | ||   |||   | | |  ||  |       | |||||   ||  |
Db        451 ATATAATATATATAATATATAATATATATTATATATAATATATTATATATTATATA 392

Qy        596 CTTATATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTA 655
              ||||||   | |  ||| | |   || |       |  | |||  |   || |
Db        391 TTTATATATTTATATATTGTATATTTATATATTGTATATATTTATATATTATATATTTAT 332

Qy        656 ATATTGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCC 715
              |||||           ||    |  ||| | | |       ||    ||    | ||
Db        331 ATATTTTTATATATTATATATATTTATATATATTATATATATTTTTATATATATTTTTAT 272

Qy        716 ATGGTATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGAC 775
              ||  |||      ||  |  || ||| |||       | | | ||     ||||||  |
Db        271 ATATTATATATTTATATATATTTTTATATATTTTTATATATTTTATATATTTATATATAT 212

Qy        776 TTCCATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTG 835
              ||  ||   | ||   || | |     ||  || | |||||    | ||||||  ||||
```



```
Db        211 TTTTATATATTATATATATTTATATATATTTTTATATATTATATATATTTATATATATTT 152

Qy        836 GTGTGATAGGAGCATTGGTTTATCATTCTATATTTTTGTTTTTTTTAGGAGAATATCTT 895

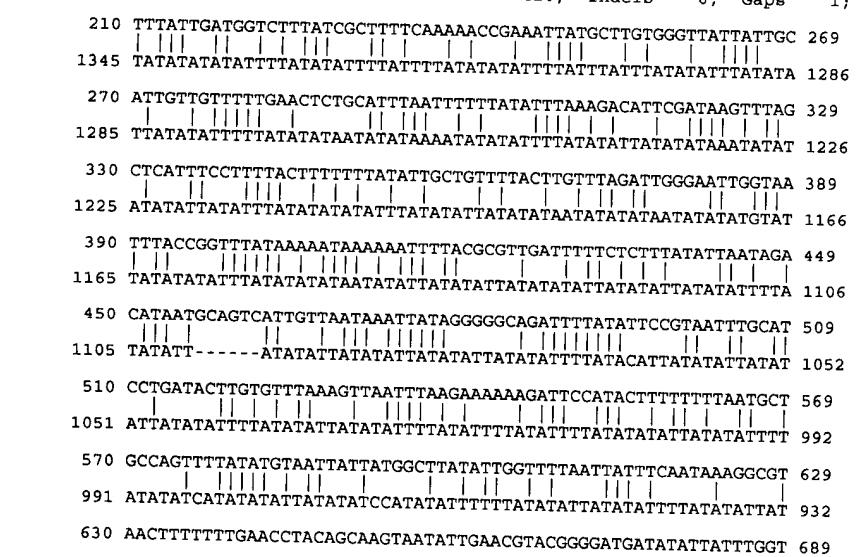
Db        151 ATATATTATATATATTTTTATATATTTTATATATTACATATATATTTATATATATTTTT 92

Qy        896 TCTTATTATAT 906

Db         91 ATATATTATAT 81
```

RESULT 5
US-09-949-016-12776/c
; Sequence 12776, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12776
; LENGTH: 187169
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(187169)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12776

Query Match 6.9%; Score 73; DB 3; Length 187169;
Best Local Similarity 44.5%; Pred. No. 1.6e-05;
Matches 341; Conservative 0; Mismatches 420; Indels 6; Gaps 1;



```
Qy        210 TTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAAATTATGCTTGTGGGTTATTATTGC 269

Db       1345 TATATATATATTTTATATATTTTATTTTATATATATTTTATTTATTTATATATTTATATA 1286

Qy        270 ATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTAAAGACATTCGATAAGTTTAG 329

Db       1285 TTATATATTTTTATATATAATATATAAAATATATATTTTATATATTATATATAAATATAT 1226

Qy        330 CTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACTTGTTTAGATTGGGAATTGGTAA 389

Db       1225 ATATATTATATTTATATATATATTTATATATTATATATAATATATATAATATATATGTAT 1166

Qy        390 TTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGATTTTTCTCTTTATATTAATAGA 449

Db       1165 TATATATATTTATATATATAATATATTATATATTATATATATTATATATTATATATTTTA 1106

Qy        450 CATAATGCAGTCATTGTTAATAAATTATAGGGGGCAGATTTTATATTCCGTAATTTGCAT 509

Db       1105 TATATT------ATATATTATATATTATATATTATATATTTTATACATTATATATTATAT 1052

Qy        510 CCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGATTCCATACTTTTTTTTAATGCT 569

Db       1051 ATTATATATTTTATATATTATATATTTTATATTTATATTTTATATATATTATATATTTT 992

Qy        570 GCCAGTTTTATATGTAATTATTATGGCTTATATTGGTTTTAATTATTTCAATAAAGGCGT 629

Db        991 ATATATCATATATATTATATATCCATATATTTTTTATATATTATATATTTTATATATTAT 932

Qy        630 AACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTACGGGGATGATATATTATTTGGT 689

Db        931 ATATTATATATATTATATATTTTATATATATTATATATTTTATATATTTTATATATTTTA 872

Qy        690 TTCACAGCTTGGTGATTATATATTCCATGGTATGGGGACATTAAATTTCTTAAATAACGG 749

Db        871 TATATATTTTATATATTTTATATATTATATATATTTTATATATTATATATTTTATATATT 812

Qy        750 CGGACAATATAAGACGTTATATGGACTTCCATCATTAATTCCTAATGACCCTCATGATTT 809

Db        811 TTATATATTTTATATATTATATATAATATATATTTTATATTTTATATATATTATATATTT 752

Qy        810 TTTATTACGGTTCTTTATAAGTATTGGTGTGATAGGAGCATTGGTTTATCATTCTATATT 869

Db        751 TATATTTTTTATATATATTATATATTTTATATTTTTATATATATTATATATTTATATATA 692

Qy        870 TTTTGTTTTTTTTAGGAGAATATCTTTCTTATTATATGAGAGAAATGCTCCTTTCATTGT 929

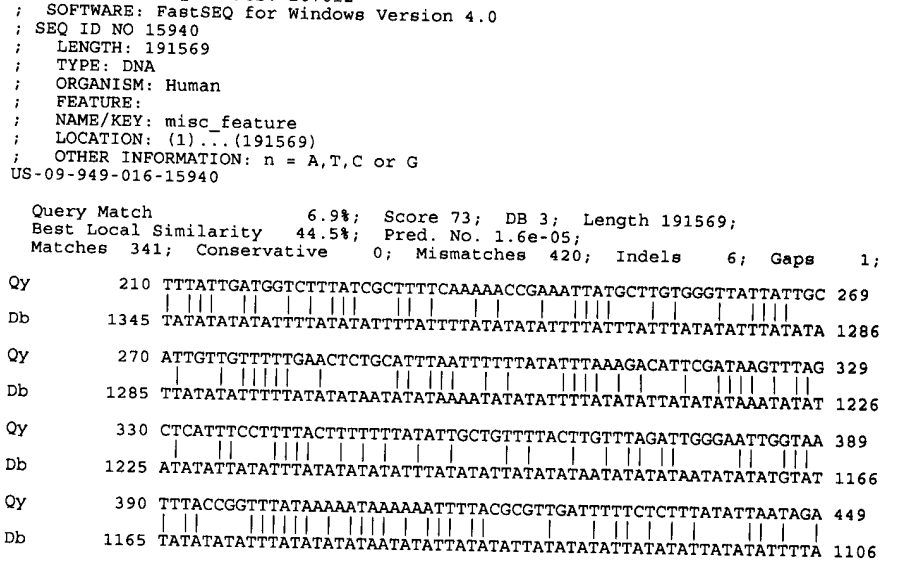
Db        691 TTTTATGTATATTATATATTATATATATTTTATATATTATATATTATATATATTTTATAT 632

Qy        930 TGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAATTTATACATTAA 976

Db        631 ATTATATATTATATATATTTATATATATTATATAGTATATATATTTA 585
```

RESULT 6
US-09-949-016-15940/c
; Sequence 15940, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15940
; LENGTH: 191569
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(191569)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15940

Query Match 6.9%; Score 73; DB 3; Length 191569;
Best Local Similarity 44.5%; Pred. No. 1.6e-05;
Matches 341; Conservative 0; Mismatches 420; Indels 6; Gaps 1;



```
Qy        210 TTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAAATTATGCTTGTGGGTTATTATTGC 269

Db       1345 TATATATATATTTTATATATTTTATTTTATATATATTTTATTTATTTATATATTTATATA 1286

Qy        270 ATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTAAAGACATTCGATAAGTTTAG 329

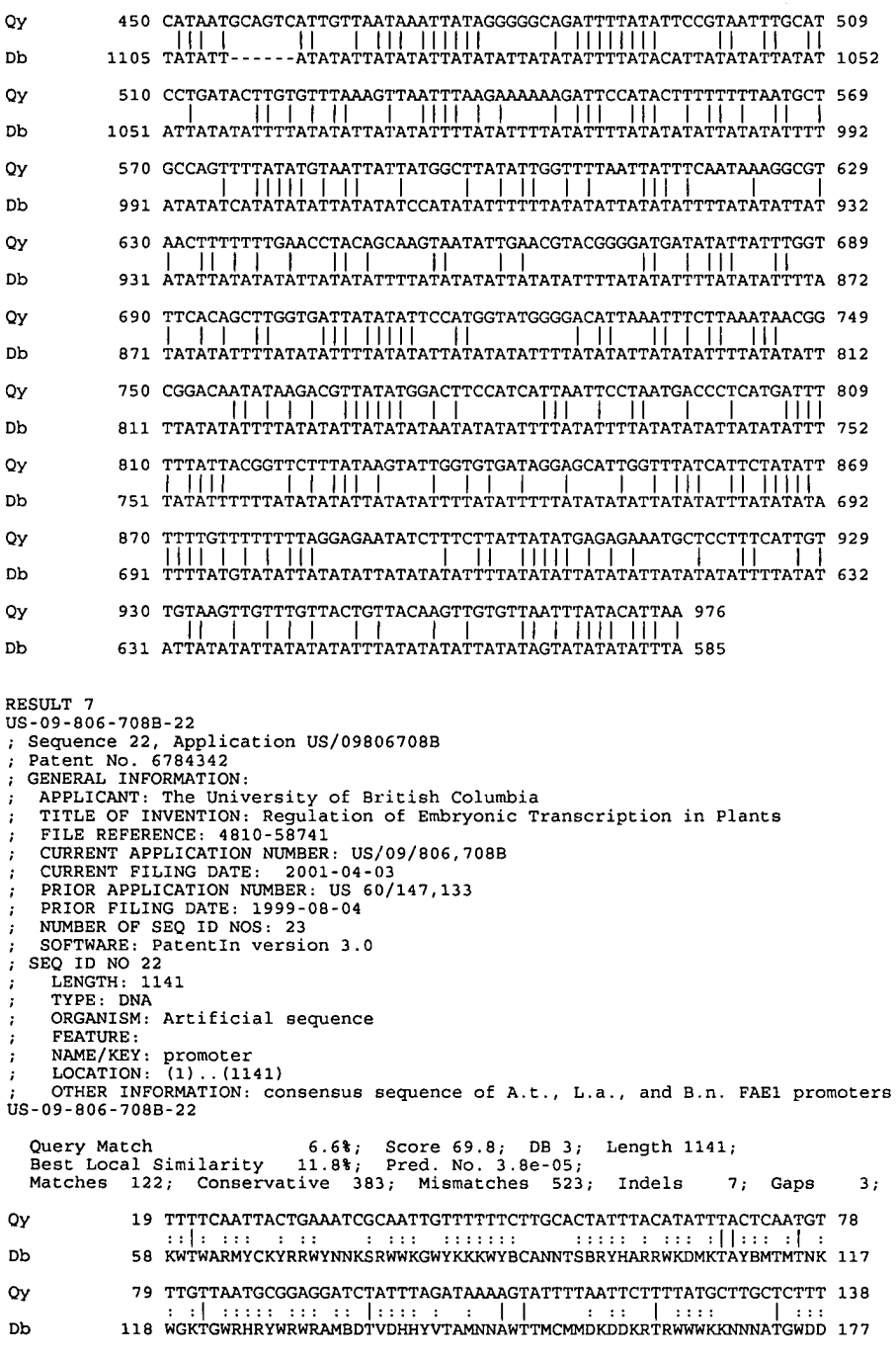
Db       1285 TTATATATTTTTATATATAATATATAAAATATATATTTTATATATTATATATAAATATAT 1226

Qy        330 CTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACTTGTTTAGATTGGGAATTGGTAA 389

Db       1225 ATATATTATATTTATATATATATTTATATATTATATATAATATATATAATATATATGTAT 1166

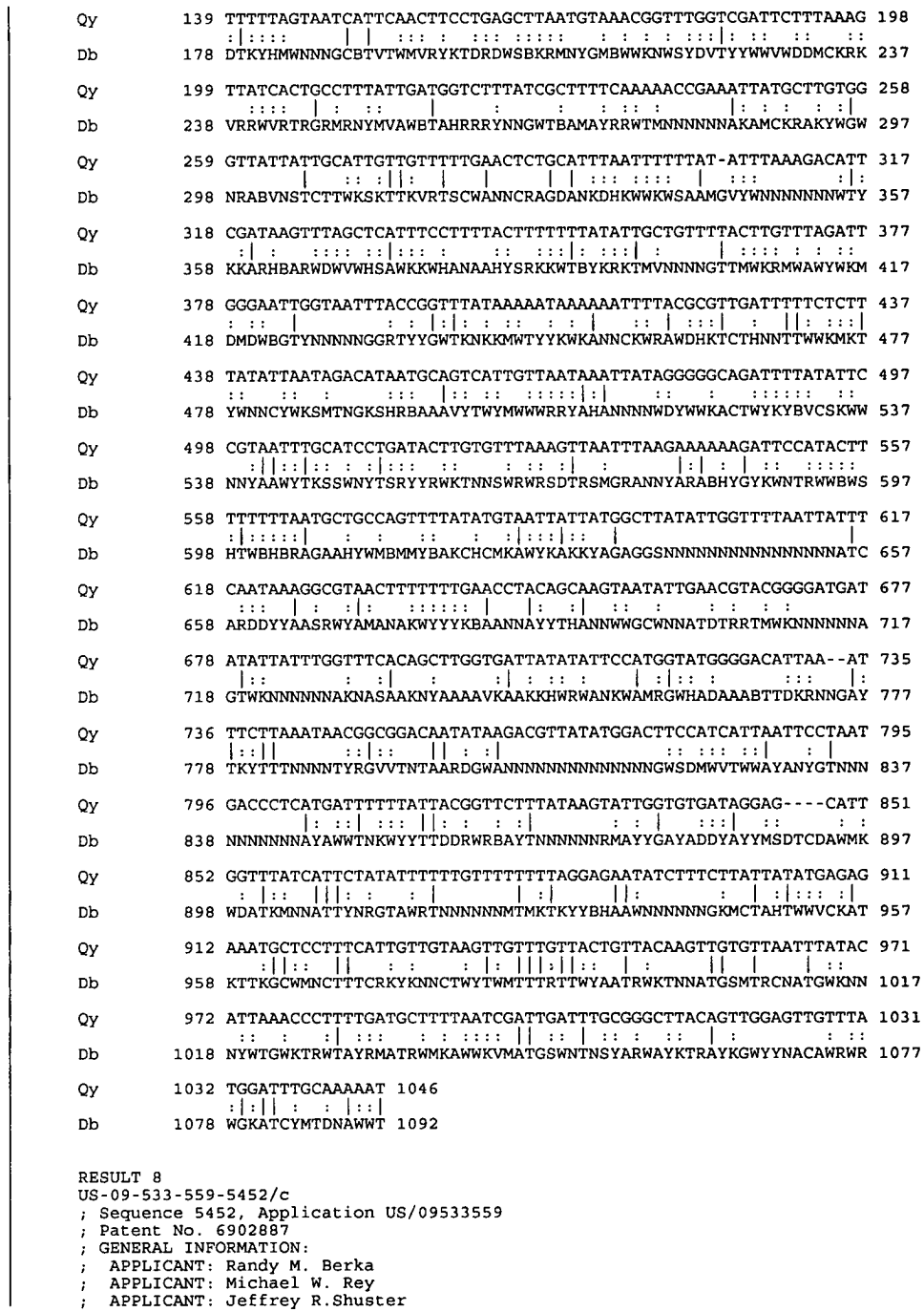
Qy        390 TTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGATTTTTCTCTTTATATTAATAGA 449

Db       1165 TATATATATTTATATATATAATATATTATATATTATATATATTATATATTATATATTTTA 1106
```



RESULT 7
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

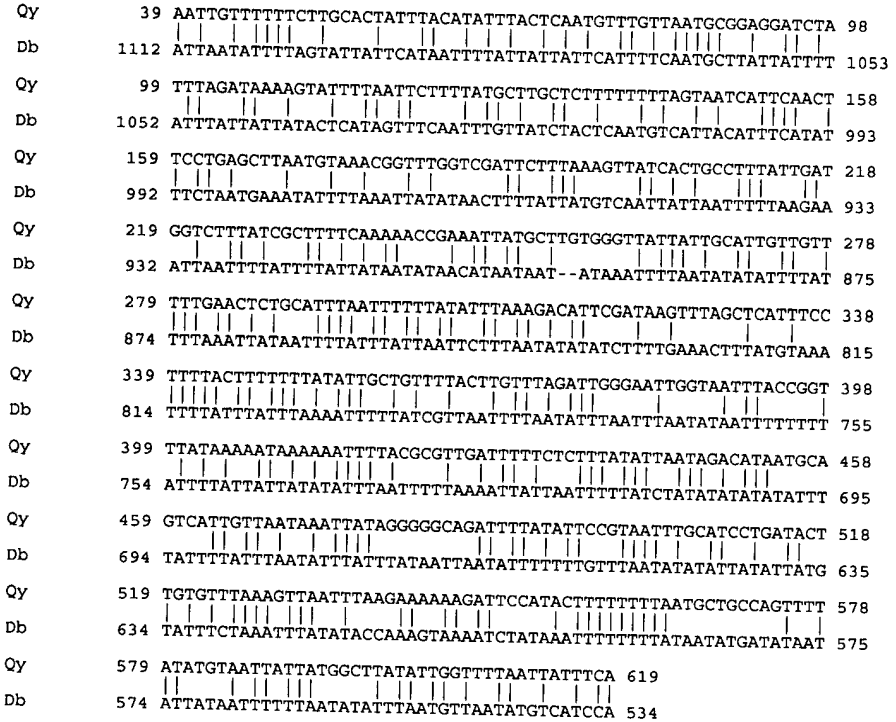
Query Match 6.6%; Score 69.8; DB 3; Length 1141;
Best Local Similarity 11.8%; Pred. No. 3.8e-05;
Matches 122; Conservative 383; Mismatches 523; Indels 7; Gaps 3;



RESULT 8
US-09-533-559-5452/c
; Sequence 5452, Application US/09533559
; Patent No. 6902887
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Jeffrey R.Shuster

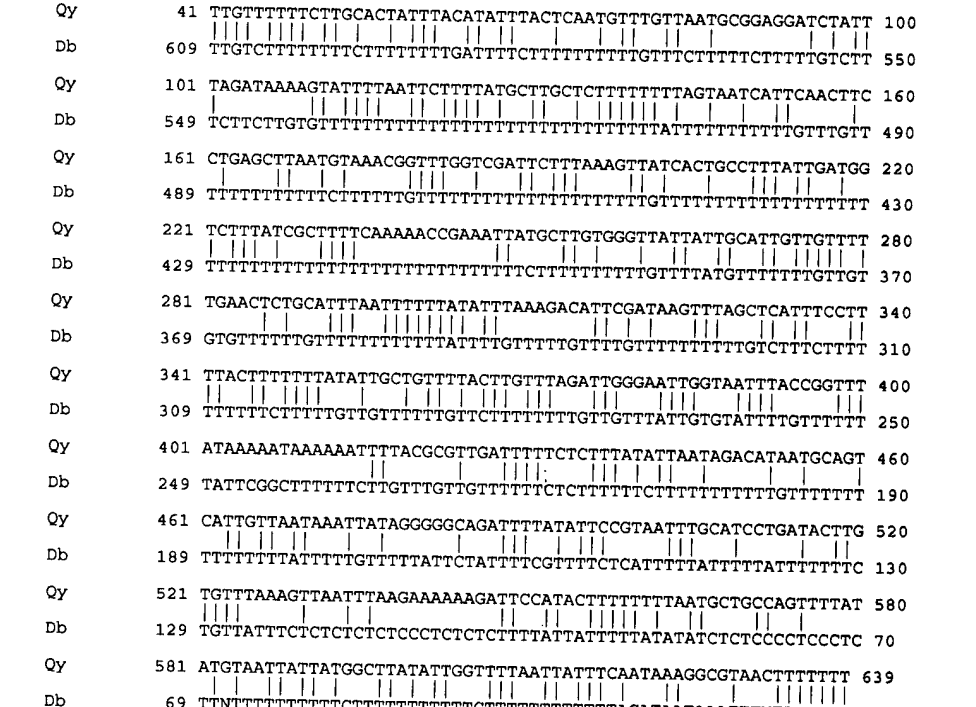
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjarke Olsen
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 5849.200-US
; CURRENT APPLICATION NUMBER: US/09/533,559
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 09/273,623
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5452
; LENGTH: 1132
; TYPE: DNA
; ORGANISM: Aspergillus oryzae
US-09-533-559-5452

Query Match 6.4%; Score 67.8; DB 3; Length 1132;
Best Local Similarity 46.0%; Pred. No. 9.6e-05;
Matches 267; Conservative 0; Mismatches 312; Indels 2; Gaps 1;



RESULT 9
US-09-902-540-1318/c
; Sequence 1318, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1318
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(614)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1318

Query Match 6.2%; Score 65.8; DB 3; Length 614;
Best Local Similarity 44.2%; Pred. No. 0.00022;
Matches 265; Conservative 0; Mismatches 334; Indels 0; Gaps 0;



RESULT 10
US-09-949-016-12896
; Sequence 12896, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

```
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12896
;   LENGTH: 451924
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12896

  Query Match             6.1%;  Score 64.6;  DB 3;  Length 451924;
  Best Local Similarity   47.0%;  Pred. No. 0.00087;
  Matches  199;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy        260 TTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTAAAGACATTCG 319
              | | | |   ||| | |  ||  |  | |  |||||||| | ||||||||       |
Db     140316 TGACTGTGAGATTATGGAATTAAATATATATATTTAATTATATATATTTATTATATATAA 140375

Qy        320 ATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACTTGTTTAGATTGG 379
              ||| | ||  | |||   | ||  || | |   |||  |  | |     | || |||
Db     140376 TTAATTATATTTAATTATATATTTATTATATATAATTAATTATATTTAATTATATATTTA 140435

Qy        380 GAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGATTTTTCTCTTTA 439
              ||   ||||| |     ||  |  |   |  ||| ||     | | || | | |
Db     140436 TTATATATAATTAATTATATTTAATTATATATTTATTATATATAATTAATTATATTTAAT 140495

Qy        440 TATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCAGATTTTATATTCCG 499
              |||  ||  |    ||  | | | |  || | |||||||         || |   |||
Db     140496 TATATATTTATTATATATAATTAATTATATTTAATTATATATTTATTATATATAATTAAT 140555

Qy        500 TAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGATTCCATACTTTT 559
              ||  ||  ||  |     |  | | || | |||||  |   ||   |||  |||   ||
Db     140556 TATATTTAATTATATATTTATTATATATAATTAATTATATTTAATTAATTATATATAATT 140615

Qy        560 TTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGGTTTTAATTATTTCA 619
              | ||    |     | |     | | |  ||||  | ||  |   |     ||| | |
Db     140616 ATATATAATTATATATAATTATATTTAATTATATGTTTAATTATATATAATTATATATAA 140675

Qy        620 ATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTACGGGGATGATAT 679
              || | |  ||| | | | | |   || |   | | ||||  ||          || ||||
Db     140676 TTATATGTTTAATTATATATAATTATATATTTAATTATATATAATTATATATAATTATAT 140735

Qy        680 ATT 682
              |||
Db     140736 ATT 140738


RESULT 11
US-09-949-016-17305
; Sequence 17305, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17305
;   LENGTH: 451925
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17305

  Query Match             6.1%;  Score 64.6;  DB 3;  Length 451925;
  Best Local Similarity   47.0%;  Pred. No. 0.00087;
  Matches  199;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy        260 TTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTAAAGACATTCG 319
              | | | |   ||| | |  ||  |  | |  |||||||| | ||||||||       |
Db     140316 TGACTGTGAGATTATGGAATTAAATATATATATTTAATTATATATATTTATTATATATAA 140375

Qy        320 ATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACTTGTTTAGATTGG 379
              ||| | ||  | |||   | ||  || | |   |||  |  | |     | || |||
Db     140376 TTAATTATATTTAATTATATATTTATTATATATAATTAATTATATTTAATTATATATTTA 140435

Qy        380 GAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGATTTTTCTCTTTA 439
              ||   ||||| |     ||  |  |   |  ||| ||     | | || | | |
Db     140436 TTATATATAATTAATTATATTTAATTATATATTTATTATATATAATTAATTATATTTAAT 140495

Qy        440 TATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCAGATTTTATATTCCG 499
              |||  ||  |    ||  | | | |  || | |||||||         || |   |||
Db     140496 TATATATTTATTATATATAATTAATTATATTTAATTATATATTTATTATATATAATTAAT 140555

Qy        500 TAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGATTCCATACTTTT 559
              ||  ||  ||  |     |  | | || | |||||  |   ||   |||  |||   ||
Db     140556 TATATTTAATTATATATTTATTATATATAATTAATTATATTTAATTAATTATATATAATT 140615

Qy        560 TTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGGTTTTAATTATTTCA 619
              | ||    |     | |     | | |  ||||  | ||  |   |     ||| | |
Db     140616 ATATATAATTATATATAATTATATTTAATTATATGTTTAATTATATATAATTATATATAA 140675

Qy        620 ATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTACGGGGATGATAT 679
              || | |  ||| | | | | |   || |   | | ||||  ||          || ||||
Db     140676 TTATATGTTTAATTATATATAATTATATATTTAATTATATATAATTATATATAATTATAT 140735

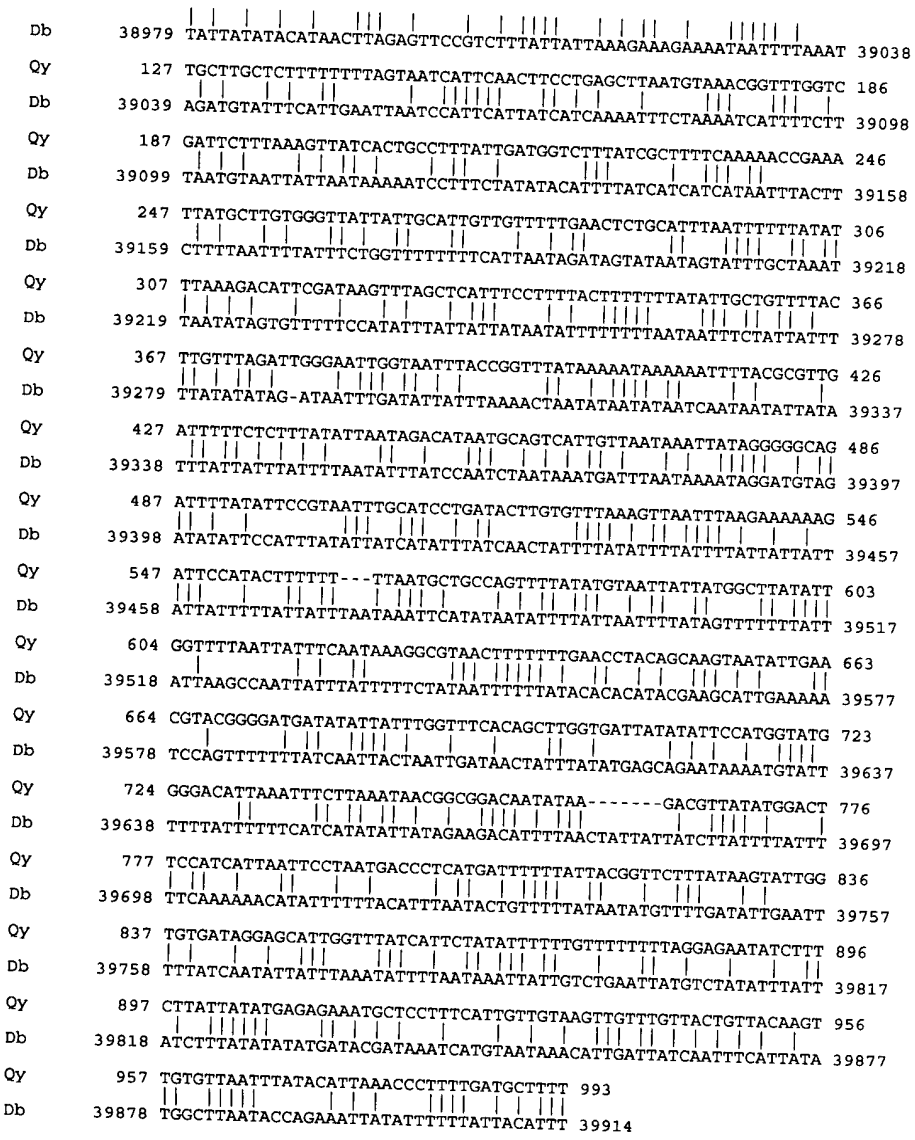
Qy        680 ATT 682
              |||
Db     140736 ATT 140738


RESULT 12
US-09-662-254B-24
; Sequence 24, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-24

  Query Match             6.0%;  Score 63.4;  DB 3;  Length 50000;
  Best Local Similarity   43.6%;  Pred. No. 0.0011;
  Matches  435;  Conservative    0;  Mismatches  551;  Indels   11;  Gaps    3;

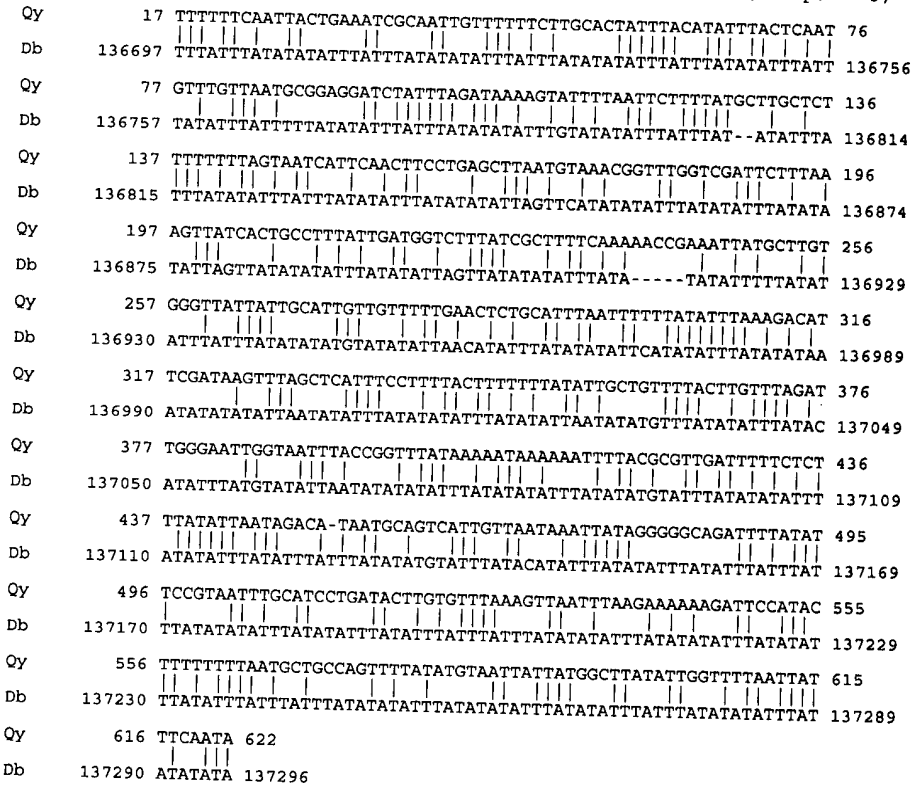
Qy          7 ATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTACATA 66
              ||| ||| | |  ||      || | |  |    |    |||||| ||     |    | |||
Db      38919 ATTTATTAATTAGTTCTTTATATTCATTTAATATAATTTTTTTATTTTCTTCCGAAGATA 38978

Qy         67 TTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTTA 126
```



RESULT 13
US-09-949-016-14624
; Sequence 14624, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14624
; LENGTH: 147382
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(147382)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14624

Query Match 6.0%; Score 63.4; DB 3; Length 147382;
Best Local Similarity 47.4%; Pred. No. 0.0013;
Matches 288; Conservative 0; Mismatches 311; Indels 8; Gaps 3;



RESULT 14
US-09-662-254B-23
; Sequence 23, Application US/09662254B
; Patent No. 6933145

```
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-23

  Query Match             6.0%;  Score 63;  DB 3;  Length 50000;
  Best Local Similarity   44.2%;  Pred. No. 0.0014;
  Matches  442;  Conservative    0;  Mismatches  550;  Indels    9;  Gaps    4;

Qy          6 TATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTACAT 65
              | || ||| |  | |   | | |     |   | | || ||  ||    | | |
Db      17320 TTTTAATTCATATATAATATTAATATTTTTTTTTAATATTCTTATTTTATATTTCTTAGA 17379

Qy         66 ATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTT 125
              ||  ||    ||  ||  |  |    | ||| ||||||||  |   |||   |  | |  |
Db      17380 ATATATATTATTCTTTATTTTATTTATGATATATTTAGAATATTTTATAAAATATTTACT 17439

Qy        126 ATGCTTGCTCTTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGT 185
              ||   |  | ||  | |||  ||  |||  |  |     | |  ||  || |  |    ||
Db      17440 ATATGTTATATTGTCTTTTATATACATATATTGTTTATTGTTATATTTATATTGCAATGT 17499

Qy        186 CGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAA 245
              ||   ||  ||  |||   |    |  ||     | |||||   ||    |||
Db      17500 AGACATTTCCAATATATTTTTTATATATTTATCCATTTTTATATATTAATTTAAAATTTC 17559

Qy        246 ATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATA 305
              |  |   ||       | | |||   |||  |  |||||   | |   | | || |   | |
Db      17560 AAAAACATTTATTTGTCTAATGTTATTAATAATTTTG---TATATTTATTATATAACAAA 17616

Qy        306 TTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTA 365
              | |||| |   |  | | |  |  ||  ||||   |||   || |  | || |  | | |
Db      17617 TATAAATATTCTGAAGATATAAAGTGCATTGAATTTGCGTTGTAATCTAATTTTTTATAT 17676

Qy        366 CTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTT 425
              || |||  |||   |  |    ||  |    |||   ||||||   |||   |  | ||
Db      17677 TTTTTTTTAATTTCTATATCTGGATAAAATTTTTTTATAAAATATTTAATGAAAATCTTT 17736

Qy        426 GATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCA 485
              || |   | |||  ||||  ||   |||  |   |  |  |      ||||  ||  |
Db      17737 -ATCTGAATATTTTTATTTTTAATAATATAGTCATATTCTTCGTATAATTTAAGTAGAAT 17795

Qy        486 GATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAA 545
              |  ||    |  |  | |  |  |  | |  | ||||  | |  |  || | ||
Db      17796 ATTATTTATATATTTTTTATTTGTATTATAATTAATTTTTATATTATCTAAAATATAATT 17855

Qy        546 GATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGG 605
              |||  |  |  | || | |||  |     ||| |  | | ||  ||  |      ||||
Db      17856 TATTAAACATATATTGTCAATTAT----ATTTAAATTATCATATTTTTTTAAATTATTAA 17911

Qy        606 TTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACG 665
              |  | |  |  | ||  |     ||  ||| | | ||     |    ||| |||   || |
Db      17912 TCATCAAAACATAAAATATATAATACATTTCTATAAAATCTCCAACAGTTATAGATAAAG 17971

Qy        666 TACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATGGG 725
              ||     | | | |||| |||  | |      |     | | | |  |  ||     |
Db      17972 TATCTTTACGCTTTATTTTTTCATCTATTTTATCATATACTGTTTTATTAATAAATTTAT 18031

Qy        726 GACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATT 785
              || | |     |||| | |  | | |  |  ||  |  ||  | |   ||  ||
Db      18032 ATTTTTCATATAACATAATATCATCATAAATGTTTTGATTTATAATATAATCATATTTTT 18091

Qy        786 AATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATAGG 845
              ||  || | |   |  | | |  |  | |     | |   |  ||  | | | |
Db      18092 TTTTATTATTTATTATTTTTAATAATATTGAATAACATATTATATATTTATTATTAGATA 18151

Qy        846 AGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTATTATA 905
              |  ||     | | | ||  |||| |   |||  |||||  | | ||| | | || | ||
Db      18152 ATGATATTAATCT-AATCATTATTATAAATTTATTTTAATATATTATATATTTTTTCATG 18210

Qy        906 TGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAAT 965
              | | ||||  |   |   |  |   | ||  |  ||     |   ||| |  ||
Db      18211 ATATAAAAATAATATCTATTTATTAAATAATTTCTATTTCTACATTAGTGTTTTTATATA 18270

Qy        966 TTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTT 1006
              || || || | |    ||| ||  | || |  | ||  ||
Db      18271 TTTTATATAATATATCTTTCATATTATTTAATGTTTTTATT 18311


RESULT 15
US-09-949-016-15851/c
; Sequence 15851, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15851
;   LENGTH: 205044
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(205044)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15851

  Query Match             5.9%;  Score 62.2;  DB 3;  Length 205044;
  Best Local Similarity   45.9%;  Pred. No. 0.0024;
  Matches  297;  Conservative    0;  Mismatches  338;  Indels   12;  Gaps    2;

Qy          4 TCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTAC 63
              | |||  ||||  | | |  ||   | ||||     ||  |||| | |      ||||||
Db     201067 TATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAA 201008

Qy         64 ATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTT 123
              |    |   | | | |  |        |  ||| || ||     ||| || ||  ||
Db     201007 AATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTT 200948

Qy        124 TTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTG 183
              |   |   | ||||| |   |    ||| ||  |     |  ||  | || |    |
Db     200947 AAAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATAT 200888

Qy        184 GTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCG 243
              |  | | | || | || |   |  | ||||  |   | || |||   ||| | |
Db     200887 TTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTAT 200828

Qy        244 AAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTA 303
              |  | |     | |  || ||||    || |  ||     |    | |  |||||  ||
Db     200827 ATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATAAATATATTTTATA 200768
```



Search completed: May 11, 2006, 05:23:27
Job time : 151.879 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:01:11 ; Search time 695.011 Seconds
(without alignments)
12528.801 Million cell updates/sec

Title: US-09-423-093B-1_COPY_9901_10953
Perfect score: 1053
Sequence: 1 atgtctattgatttactttt.....gatttgcaaaaattagataa 1053

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1053	100.0	14516	3	US-09-423-093A-1	Sequence 1, Appli
c	2	123.4	11.7	1300	7	US-10-668-749A-1	Sequence 1, Appli
	3	116.6	11.1	15387	6	US-10-311-455-157	Sequence 157, App
	4	113.2	10.8	7442	7	US-10-221-714A-409	Sequence 409, App
	5	107.6	10.2	7814	7	US-10-221-714A-252	Sequence 252, App
	6	107.4	10.2	7597	6	US-10-311-455-986	Sequence 986, App
	7	107	10.2	6668	6	US-10-311-455-1670	Sequence 1670, Ap
	8	105.6	10.0	9539	5	US-10-239-676-52	Sequence 52, Appl
	9	105.6	10.0	9539	6	US-10-240-453-54	Sequence 54, Appl
	10	102.2	9.7	6292	7	US-10-221-714A-461	Sequence 461, App
	11	102	9.7	3673778	6	US-10-312-841-2	Sequence 2, Appli
	12	101.8	9.7	13511	6	US-10-311-455-254	Sequence 254, App
	13	100.6	9.6	1243	8	US-10-425-115-172717	Sequence 172717,
	14	100.4	9.5	9539	5	US-10-239-676-51	Sequence 51, Appl
	15	100.4	9.5	9539	6	US-10-240-453-53	Sequence 53, Appl
	16	100.2	9.5	8392	6	US-10-311-455-1463	Sequence 1463, Ap
	17	99	9.4	14006	6	US-10-311-455-1931	Sequence 1931, Ap
	18	98	9.3	3683	8	US-10-473-126-339	Sequence 339, App
	19	98	9.3	73334	6	US-10-311-455-2097	Sequence 2097, Ap
	20	98	9.3	73334	7	US-10-240-589C-127	Sequence 127, App
	21	96.8	9.2	1297	8	US-10-425-115-67687	Sequence 67687, A
	22	96.8	9.2	17967	6	US-10-311-455-988	Sequence 988, App
	23	96.2	9.1	6419	6	US-10-311-455-240	Sequence 240, App
	24	95.6	9.1	1062	8	US-10-425-115-120013	Sequence 120013,
	25	94.2	8.9	5413	7	US-10-221-714A-418	Sequence 418, App
	26	94	8.9	1204	7	US-10-437-963-77858	Sequence 77858, A
	27	94	8.9	3683	8	US-10-473-126-193	Sequence 193, App
	28	93.8	8.9	17594	6	US-10-311-455-1999	Sequence 1999, Ap
c	29	93.4	8.9	1214	7	US-10-424-599-102083	Sequence 102083,
	30	93.4	8.9	6259	6	US-10-311-455-263	Sequence 263, App
	31	93.4	8.9	6286	7	US-10-221-714A-313	Sequence 313, App
	32	93.2	8.9	6767	7	US-10-221-714A-330	Sequence 330, App
	33	93	8.8	15548	6	US-10-311-455-2128	Sequence 2128, Ap
	34	92.8	8.8	921	8	US-10-425-115-38710	Sequence 38710, A
	35	92.6	8.8	5893	6	US-10-311-455-832	Sequence 832, App
	36	92.2	8.8	14950	6	US-10-311-455-1230	Sequence 1230, Ap
	37	91.8	8.7	83391	7	US-10-433-793-124	Sequence 124, App
	38	91	8.6	5750	6	US-10-311-455-1981	Sequence 1981, Ap
	39	91	8.6	5750	7	US-10-221-714A-433	Sequence 433, App
	40	91	8.6	8222	9	US-10-486-319A-65	Sequence 65, Appl
c	41	89.8	8.5	8056	8	US-10-473-126-386	Sequence 386, App
	42	89.8	8.5	16766	6	US-10-311-455-2130	Sequence 2130, Ap
	43	89.6	8.5	6171	6	US-10-311-455-761	Sequence 761, App
	44	89.4	8.5	975	8	US-10-363-345A-16099	Sequence 16099, A
c	45	89.4	8.5	975	8	US-10-363-345A-16100	Sequence 16100, A

ALIGNMENTS

RESULT 1
US-09-423-093A-1
; Sequence 1, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 14516
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-423-093A-1

Query Match 100.0%; Score 1053; DB 3; Length 14516;
Best Local Similarity 100.0%; Pred. No. 3.1e-176;
Matches 1053; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

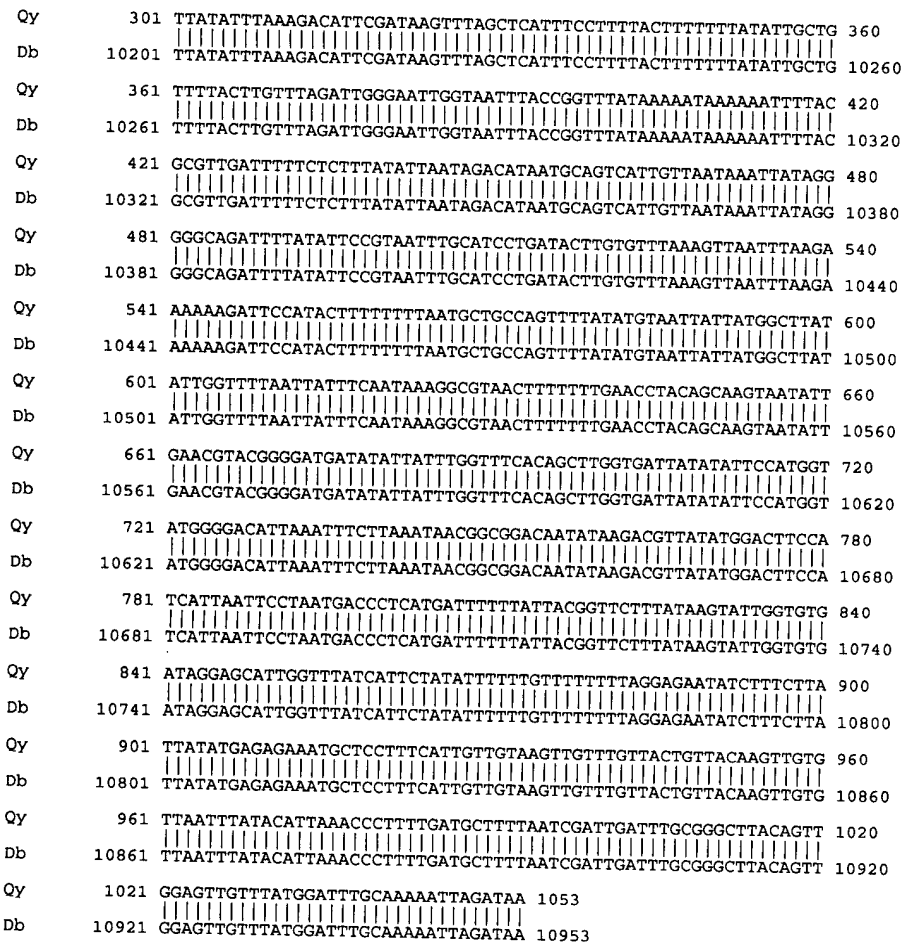
```
Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9901 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 9960

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9961 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 10020

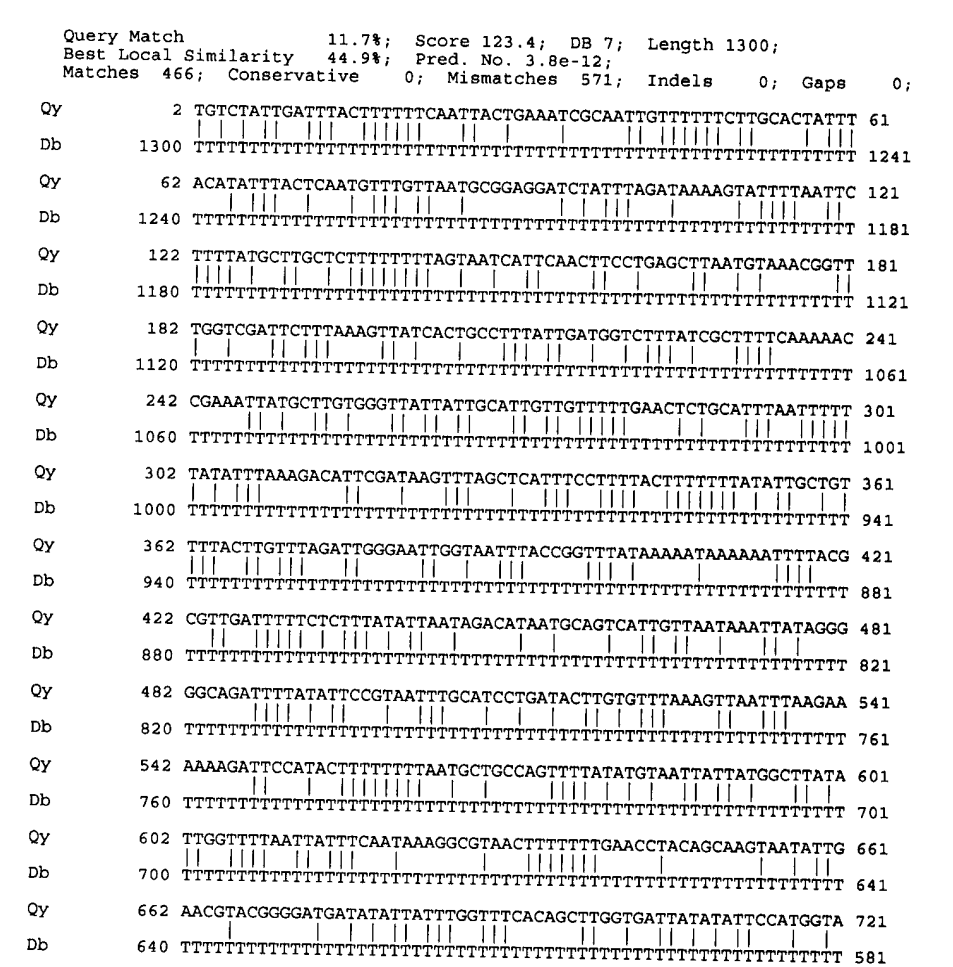
Qy        121 CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10021 CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 10080

Qy        181 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10081 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 10140

Qy        241 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10141 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 10200
```



RESULT 2
US-10-668-749A-1/c
; Sequence 1, Application US/10668749A
; Publication No. US20040110205A1
; GENERAL INFORMATION:
; APPLICANT: Agilent Technologies
; TITLE OF INVENTION: Methods and Systems for Nanopore Data Analysis
; FILE REFERENCE: 50112-1580
; CURRENT APPLICATION NUMBER: US/10/668,749A
; CURRENT FILING DATE: 2003-09-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
US-10-668-749A-1



```
Qy       1022 GAGTTGTTTATGGATTT 1038
              || ||| |   |||
Db        280 TTTTTTTTTTTTTTTT 264


RESULT 3
US-10-311-455-157
; Sequence 157, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 157
;   LENGTH: 15387
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-157

  Query Match             11.1%;  Score 116.6;  DB 6;  Length 15387;
  Best Local Similarity   47.1%;  Pred. No. 1.3e-10;
  Matches  490;  Conservative    0;  Mismatches  544;  Indels    7;  Gaps    4;

Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              || | |||||  | |  |||| | |  || | |  |    |  ||| || ||    | |
Db      10087 ATTTTTATTGTTATTTTTTTATTAGATATTAAGTTTTGTAAATTTTATTTTTTATTTTTG 10146

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
              |  || |||| |     ||| ||| |    |  |  | ||   |       | ||||| ||
Db      10147 TTTATTTTTATTGTTATTTTTTTATTAGATATTAAGTTTTGTAAATTTTATTTTTTATTT 10206

Qy        121 ---CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAAC 177
                 || | |   |  |  |||||||| ||   ||| |  ||  | |  || || |
Db      10207 TTGTTTGTTTTTATTGTTATTTTTTTATTAGATATTAAGTTTTGTAAATTTTATTTTTTA 10266

Qy        178 GGTTTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAA 237
              ||||  | |  | | |     | |    |       | | | | | ||    ||
Db      10267 TTTTTGTTTGTATTTATTGTTATTTTTTTATTAGATATTAAGTTTTGTAAATTTTAATTT 10326

Qy        238 AAACCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAAT 297
              |      ||    ||    ||||||  ||  |||     ||  |   | |  |||| ||
Db      10327 TTATTTTTGTTTATTTTTATTGTTATTTTTTTTATTAGATATTAAGTTTTGTAAATTTTAT 10386

Qy        298 TTTTTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTG 357
              |||||||  |||     |  ||   |    ||   | |||   | |||  |||| || |||
Db      10387 TTTTTATTTTTGTTTATTTTTATTGTTATTTTTTTATTAGATATTAAGTTTTGTAAATTT 10446

Qy        358 CTGTTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTT 417
              ||||  || ||      |    |||| || |||     |  |||  ||
Db      10447 TATTTTTTATTTTTGTTTATTTTTATTGTTATTTTTTTATTAGATATTAAGTTTTGTAAA 10506

Qy        418 TACGCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTAT 477
              |      ||  | ||| | ||||| || || |  ||  |    | |    | |    || |
Db      10507 TTTTATTTTTTATTTTTGTTTATTTTTATTGTTATTTTTTTATTAGATATTAAGTTTTGT 10566

Qy        478 AGGGGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTA 537
              |      | |||||| ||    |  |||  ||  | ||  || | ||  |  |||| ||
Db      10567 AAATTTTATTTTTTATTTTTGTTTGTTTTTATTGTTATTTTTTTATTAGATATTAAGTTT 10626

Qy        538 AGAAAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCT 597
              | |||      |   |  ||||| || ||  |    ||| | | | || |   |||
Db      10627 TGTAAATTTTATTTTTTATTTTTGTTTATTTTTATTGTTATTTTTTTTATTAGATATTAAG 10686

Qy        598 TATATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAAT 657
              | |  |   ||| ||| |||    |        ||   | || ||     |  |  |  || |
Db      10687 TTTTGTAAATTTTATTTTTTATTTTTGTTTGTTTTTATTGTTATTTTTTTATTAGATATT 10746

Qy        658 ATTGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCAT 717
              |      |||          | |  | |||| ||       ||| |  || | ||||  ||
Db      10747 AAGTTTTGTAAATTTTATTTTTTATTTTGTTTGTTTTTATTGTTATTTTTTTATTAGAT 10806

Qy        718 GGTATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTT 777
              ||   |     |||||||  |   | |      |   || |     ||||| |     |
Db      10807 ATTA--AGTTTTGTAAATTTTATTTTTTATTTTTGTTTATTTTTATTGTTATTTTTTTAT 10864

Qy        778 CCATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGT 837
              |||||  |  |  | |  | ||  ||| || ||     | |||||  |||| |
Db      10865 TAGATATTAAGT-TTTGTAAATTTTATTTTTTATTTTTGTTTATTTTTATTGTTATTTTT 10923

Qy        838 GTGATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTC 897
              |  |||       | ||| | | | |   |||||  |||| ||||     ||   | |
Db      10924 TTATTAGATATTAAGTTTTGTAAATTTTATTTTTTATTTTTGTTTATTTTTATTGTTATT 10983

Qy        898 TTATTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTT 957
              || |||| || |  ||   |  |    ||  | |    | ||||||  | ||    |||
Db      10984 TTTTTATTAGATATTAAGTTTTGTAAATTTTATTTTTTATTTTTGTTTGTTTTTATTGTT 11043

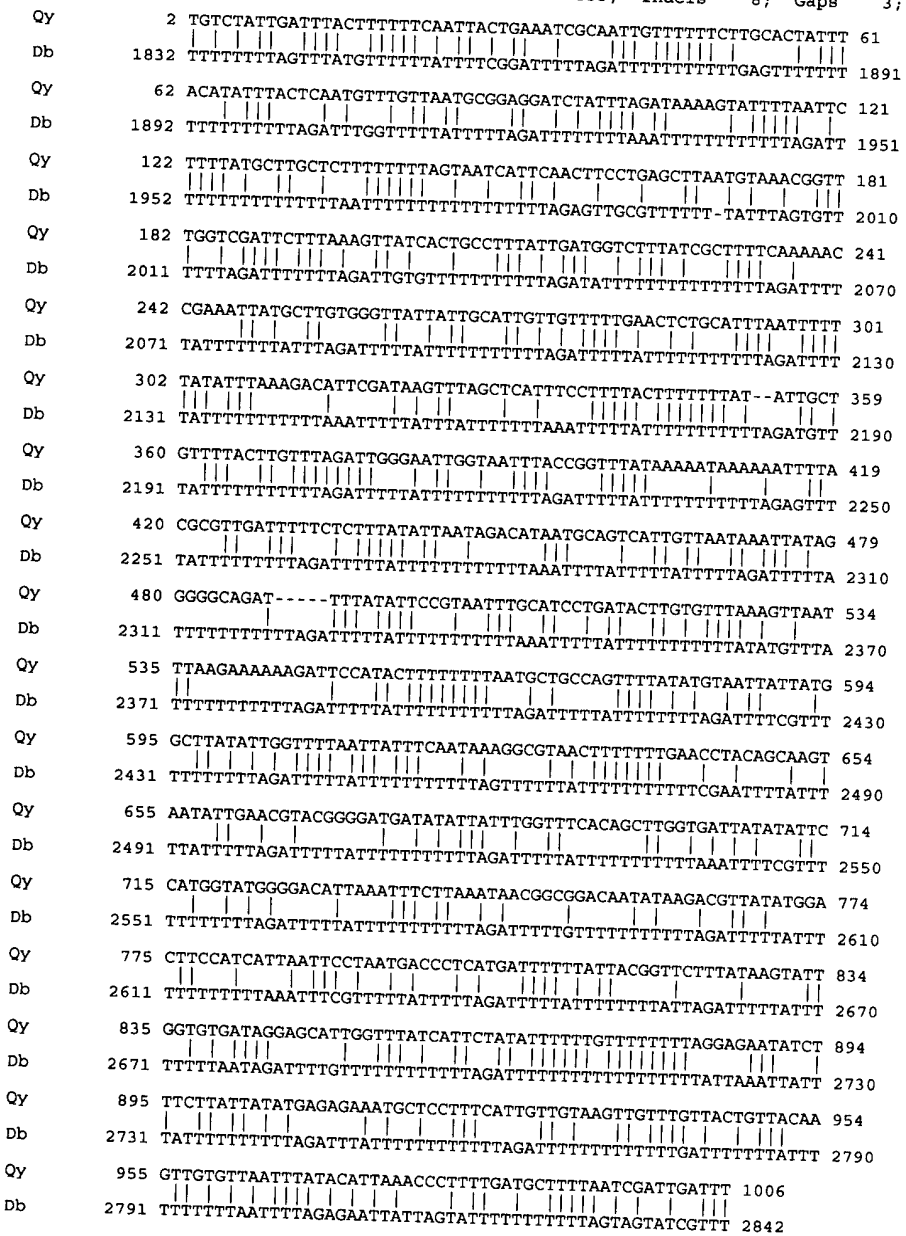
Qy        958 GTGTTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACA 1017
              | ||  | | ||   || |     |||| ||| ||||| |   ||  |  |     ||
Db      11044 ATTTTTTTATTTAGGTTTATTAATTTT-ATGTTTTTAGTTTGTTTTTAAGTTTTTTTTTT 11102

Qy       1018 GTTGGAGTTGTTTATGGATTT 1038
              ||  |||| || ||| ||||
Db      11103 TTTTAAGTTTTTGATGTATTT 11123


RESULT 4
US-10-221-714A-409
; Sequence 409, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 409
;   LENGTH: 7442
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1792, 4861..4862
;   OTHER INFORMATION: n is a or g or c or t
```

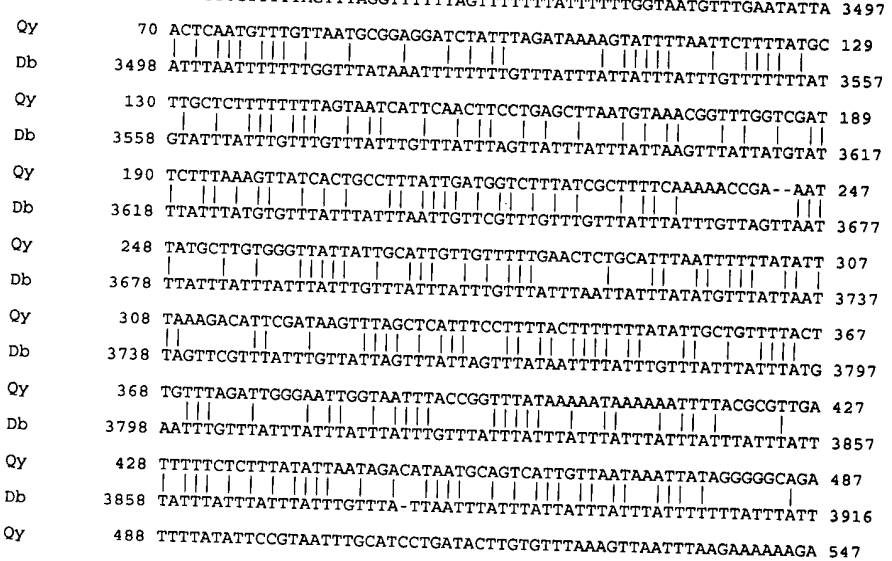
US-10-221-714A-409

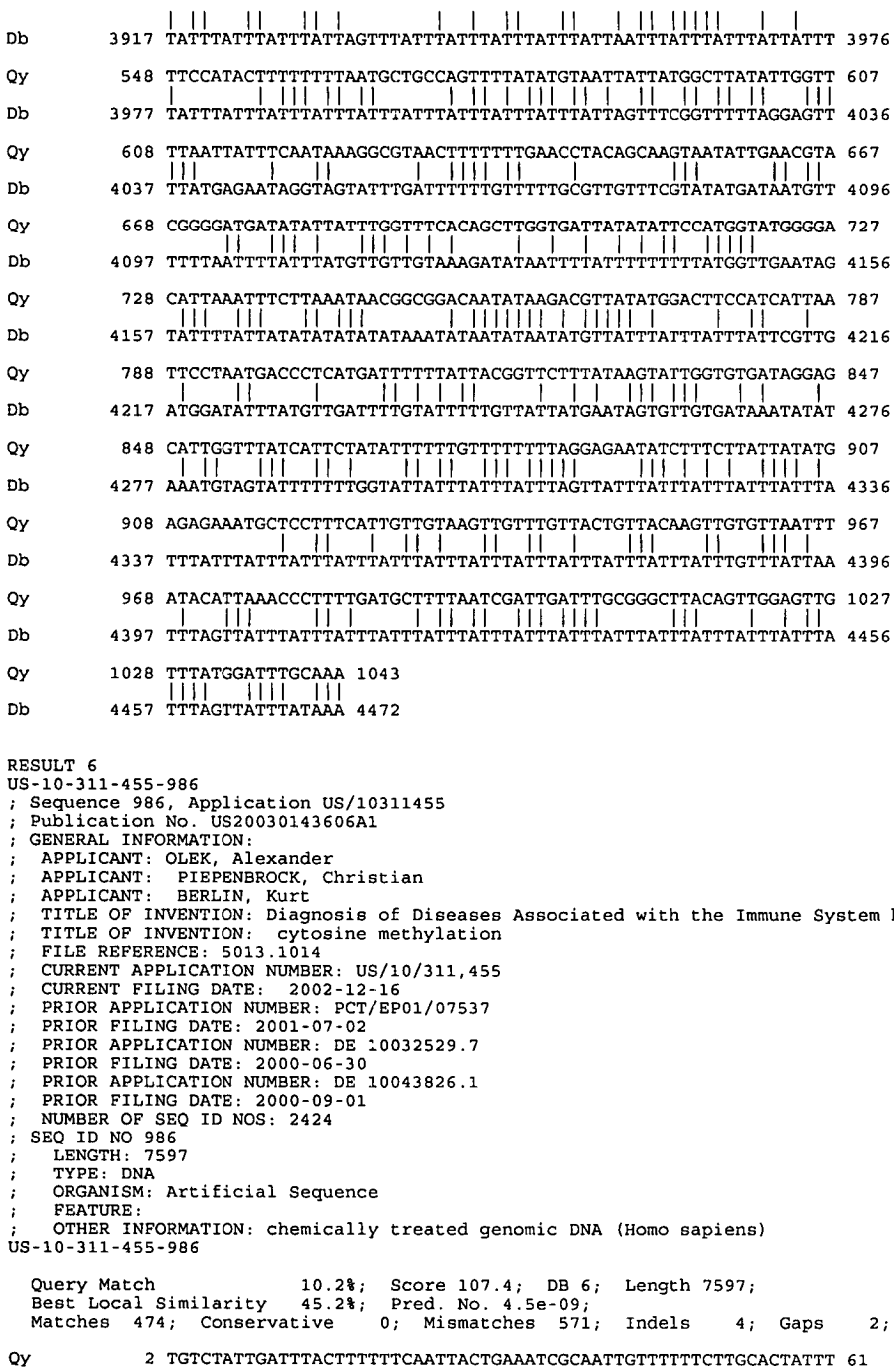
Query Match 10.8%; Score 113.2; DB 7; Length 7442;
Best Local Similarity 46.5%; Pred. No. 4.2e-10;
Matches 471; Conservative 0; Mismatches 533; Indels 8; Gaps 3;



RESULT 5
US-10-221-714A-252
; Sequence 252, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 252
; LENGTH: 7814
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-252

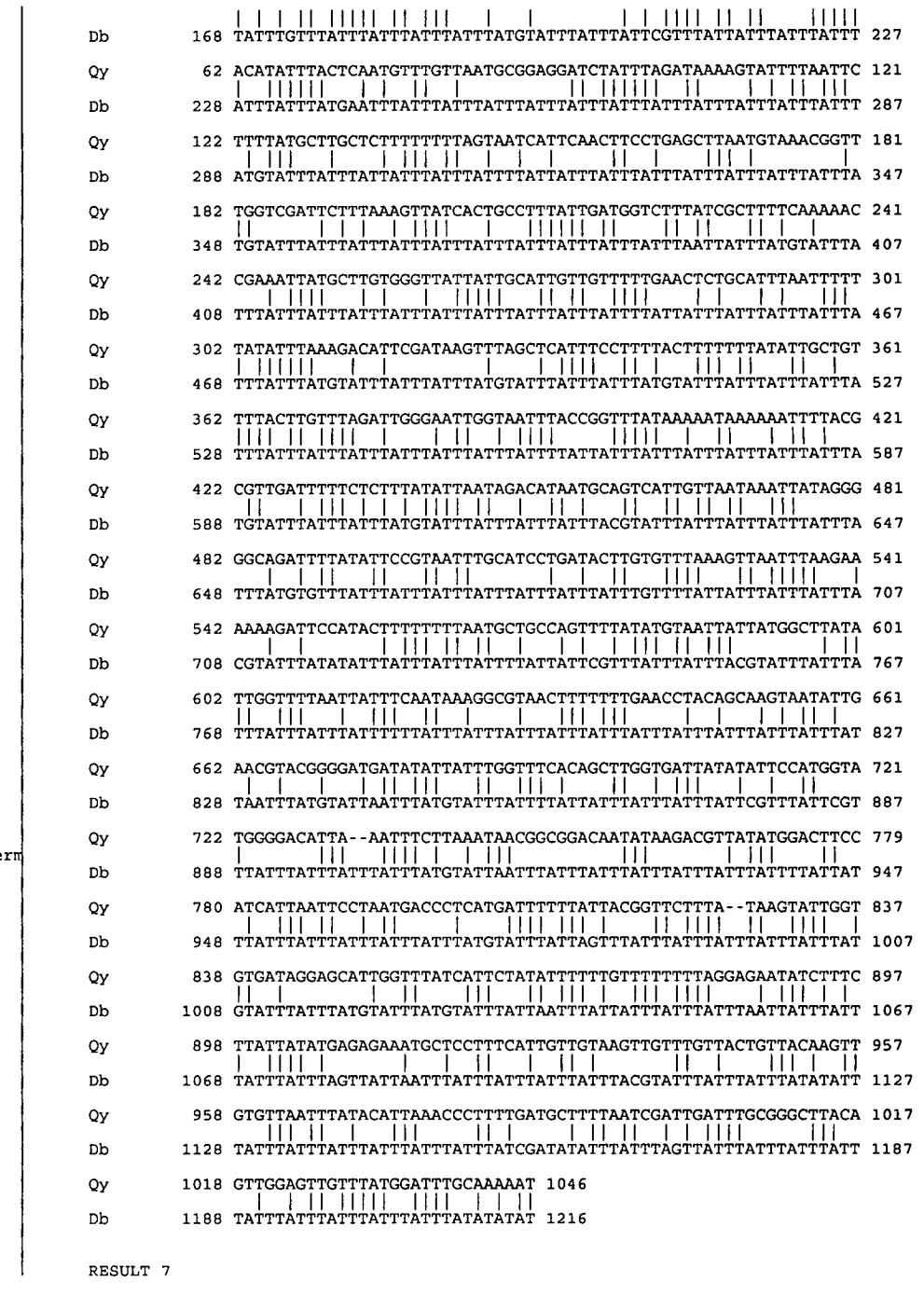
Query Match 10.2%; Score 107.6; DB 7; Length 7814;
Best Local Similarity 45.3%; Pred. No. 4.1e-09;
Matches 469; Conservative 0; Mismatches 564; Indels 3; Gaps 2;





RESULT 6
US-10-311-455-986
; Sequence 986, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 986
; LENGTH: 7597
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-986

Query Match 10.2%; Score 107.4; DB 6; Length 7597;
Best Local Similarity 45.2%; Pred. No. 4.5e-09;
Matches 474; Conservative 0; Mismatches 571; Indels 4; Gaps 2;



RESULT 7

```
US-10-311-455-1670
; Sequence 1670, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1670
;   LENGTH: 6668
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1936
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1670

  Query Match           10.2%;  Score 107;  DB 6;  Length 6668;
  Best Local Similarity   44.6%;  Pred. No. 5e-09;
  Matches  461;  Conservative    0;  Mismatches  570;  Indels    2;  Gaps    1;

Qy          6 TATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTACAT 65
Db       2752 TTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTT 2811

Qy         66 ATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTT 125
Db       2812 TTTTGTTTTTTTTTTTTTTTTTTTCGTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTGT 2871

Qy        126 ATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGT 185
Db       2872 TTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTGTTTTTTTTTGTTTTT 2931

Qy        186 CGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAA 245
Db       2932 TTTTTTTTTATTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTT 2991

Qy        246 ATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATA 305
Db       2992 TTTTTTTTTTTTTTTTTTTTTGTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3051

Qy        306 TTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTA 365
Db       3052 GTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTTTGTTTTTTT 3111

Qy        366 CTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTT 425
Db       3112 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTTTTTT 3171

Qy        426 GATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCA 485
Db       3172 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTGT 3231

Qy        486 GATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAA 545
Db       3232 TTTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTGT 3291

Qy        546 GATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGG 605
Db       3292 TTTTGTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3351

Qy        606 TTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACG 665
Db       3352 TTTTTTTTTTGTTATTTGTAGTATTAAGGGGAATAGAAGTCGTGAATTTTTAAATATTGA 3411

Qy        666 TACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATGGG 725
Db       3412 ATATTTGATTTTTTTAATTTTCGAAAAGTATGGTTTTGATTATATTTTTTTAGTTTTAGT 3471

Qy        726 GACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATT 785
Db       3472 TTTAACGAAGTTTCGTTTTTTGGAATTTGATGTATTTATTTTAAATTAAATTTTACGTTT 3531

Qy        786 --AATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATA 843
Db       3532 AGAGTTTTTAATTGTAGAGCGGTATTTGTACGGGGTTTTTTGGGGTTTTTAGTTTTTTTT 3591

Qy        844 GGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTATTA 903
Db       3592 AGTGTAGAATTTAAGGTTTTTATTTTTTTTTCGTTTTTTATTTTACGTTTATTTTTATTA 3651

Qy        904 TATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTA 963
Db       3652 GAAAATTGGGATTTTTTTTTGGGTTTTTTAGTTAGGGAGGAAATATTTTTTTTTTTATTT 3711

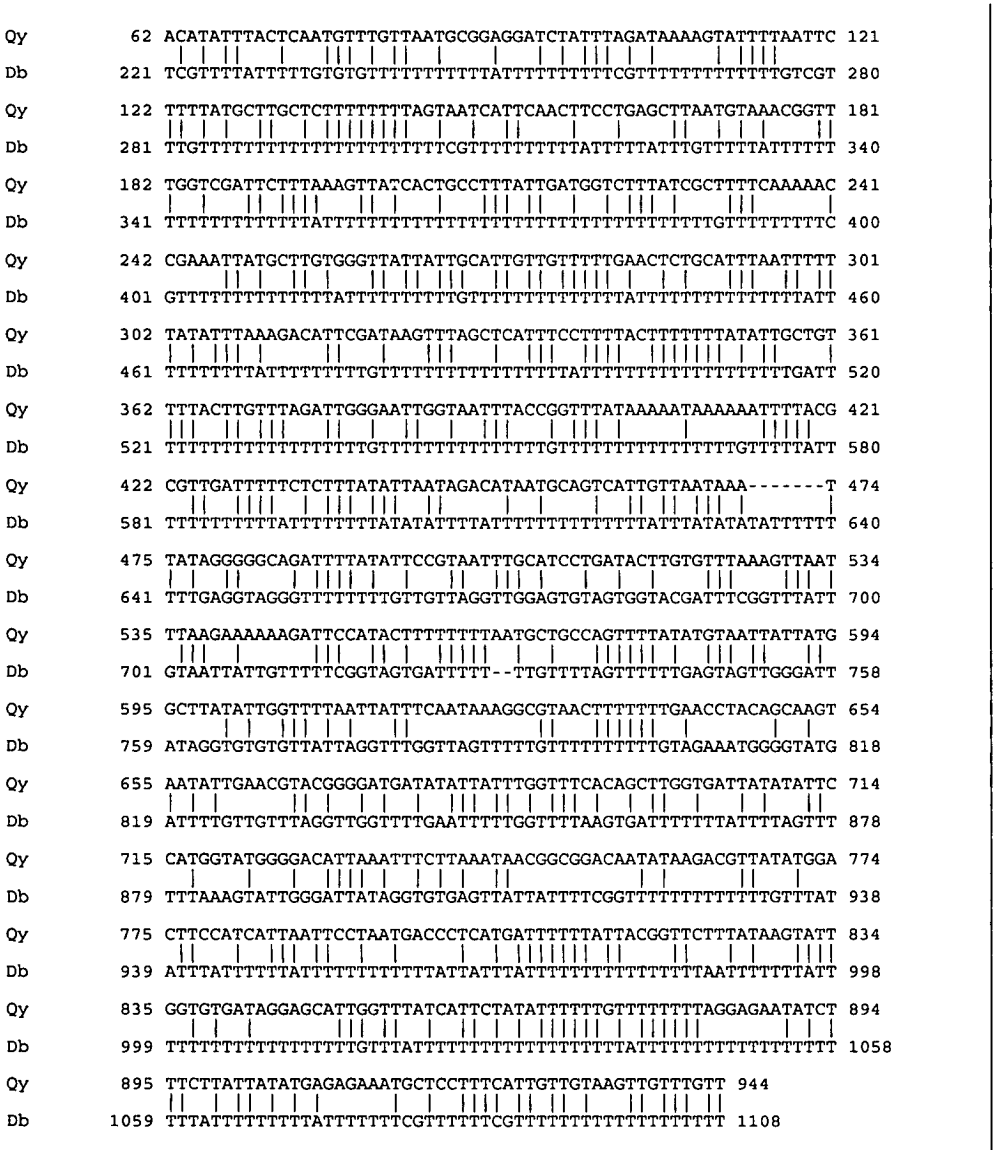
Qy        964 ATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTTGGA 1023
Db       3712 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTAGATTTTAAGTTTTAG 3771

Qy       1024 GTTGTTTATGGAT 1036
Db       3772 GGTATATGTGTAT 3784

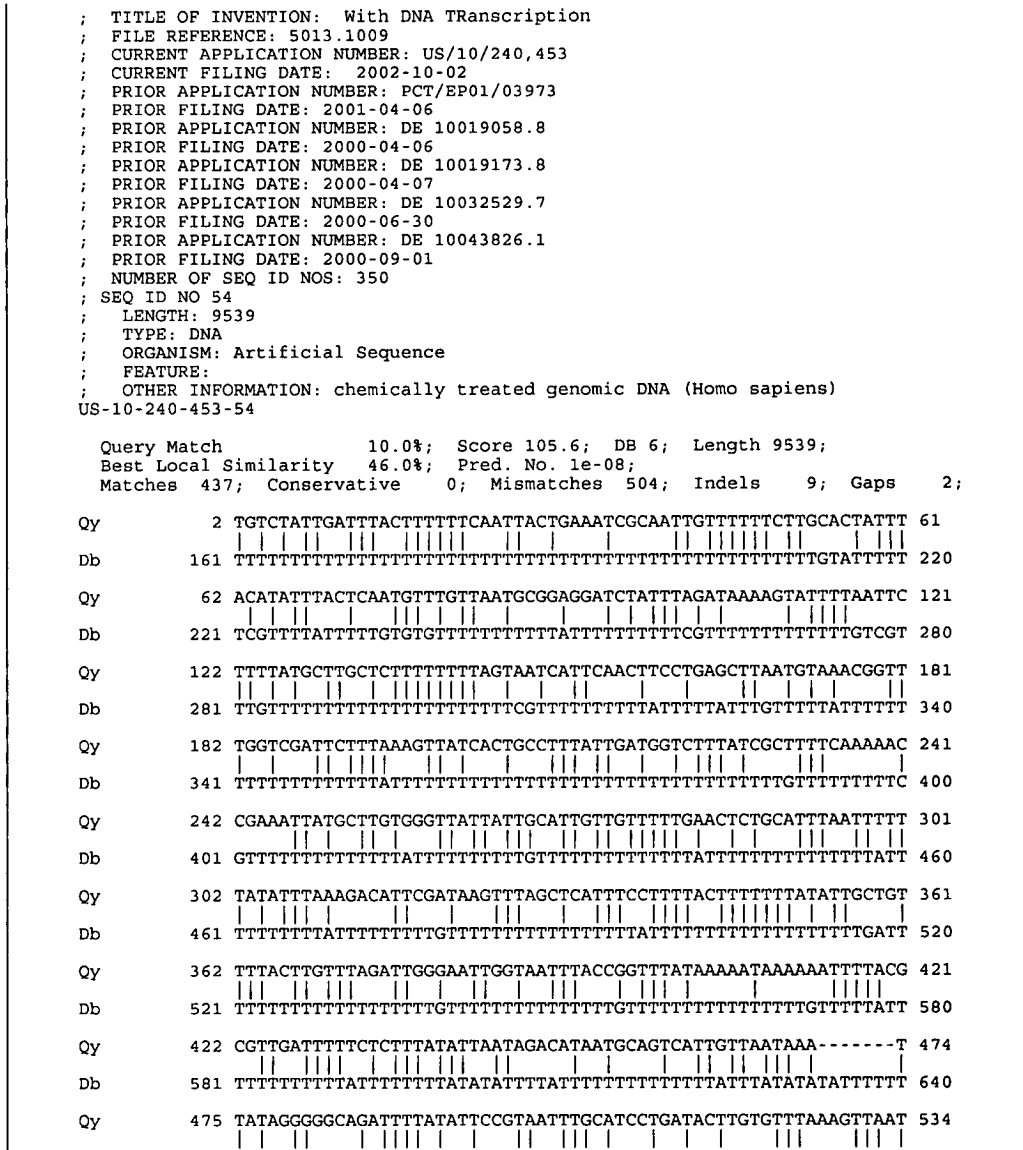

RESULT 8
US-10-239-676-52
; Sequence 52, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 52
;   LENGTH: 9539
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-52

  Query Match           10.0%;  Score 105.6;  DB 5;  Length 9539;
  Best Local Similarity   46.0%;  Pred. No. 1e-08;
  Matches  437;  Conservative    0;  Mismatches  504;  Indels    9;  Gaps    2;

Qy          2 TGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTT 61
Db        161 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTATTTTT 220
```

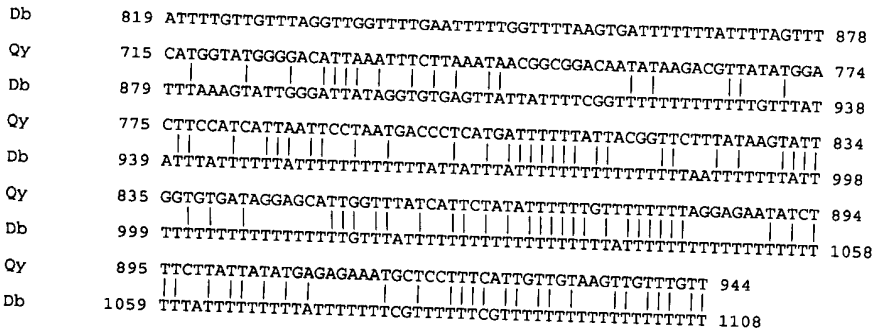


RESULT 9
US-10-240-453-54
; Sequence 54, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated



; TITLE OF INVENTION: With DNA TRanscription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 54
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-54

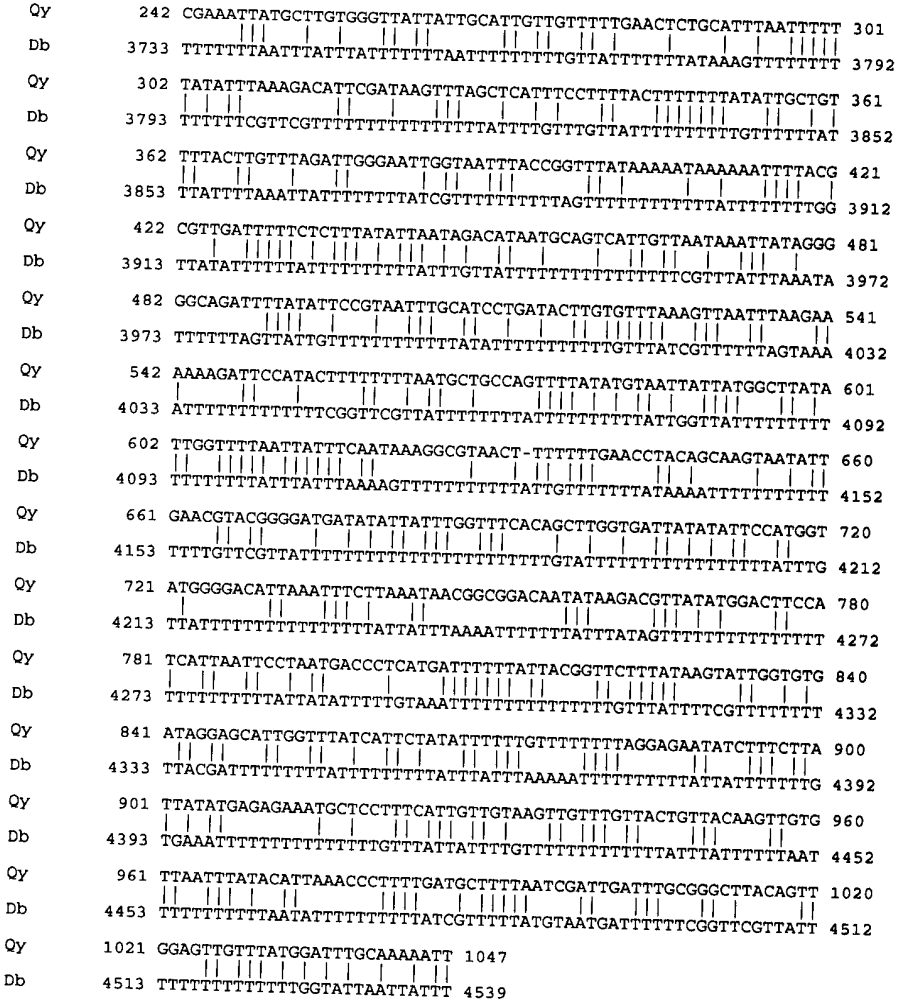
Query Match 10.0%; Score 105.6; DB 6; Length 9539;
Best Local Similarity 46.0%; Pred. No. 1e-08;
Matches 437; Conservative 0; Mismatches 504; Indels 9; Gaps 2;



RESULT 10
US-10-221-714A-461
; Sequence 461, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 461
; LENGTH: 6292
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-461

Query Match 9.7%; Score 102.2; DB 7; Length 6292;
Best Local Similarity 44.2%; Pred. No. 3.5e-08;
Matches 463; Conservative 0; Mismatches 583; Indels 1; Gaps 1;





RESULT 11
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

```
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (379615)
US-10-312-841-2

  Query Match             9.7%;  Score 102;  DB 6;  Length 3673778;
  Best Local Similarity   45.6%;  Pred. No. 2.8e-07;
  Matches  478;  Conservative    0;  Mismatches  560;  Indels   10;  Gaps    3;

Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
              || | | || | |  | | | || | ||  |    |     ||  ||||| ||    |||
Db      20421 ATTTTTTTTTAAGTTTTATATTTAGTTTTTTTTGTTTATTTTTATTTTTATTTTATATA 20480

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
              |   |||||    | | ||| ||| |       |    | |||  | |  | |||||  |
Db      20481 TTTTTATTTTTGTAGTTTTTTTTATTTTTATAAAGAGTTATAGGAATAGATTTTTTATTA 20540

Qy        121 CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 180
              ||||  | ||    ||| || |     | |||  | ||  |    ||  | ||    ||
Db      20541 TTTTTTAGTTTTAGTTTTATTATTTATTTAATTTTAATTTTTTGCGTTTGTTTAGGTTGT 20600

Qy        181 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 240
              ||  |   ||  ||  | ||||    |   |||||  ||    ||| |   ||||  |
Db      20601 TTATTTTTTTAATTTTATTTATTTTTTTTTTTATATATACGTTTTTTTTGTTTTAGATCG 20660

Qy        241 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 300
              |  ||     | |      || |||   | | ||||||  | | ||  |||| ||||
Db      20661 GTTATTTTTATTGTTTTTTAATTTGTTGGTTATTGGTTTTTTTAGTTTGTTTTTATTTTT 20720

Qy        301 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTTATATTGCTG 360
              |   |||          ||  ||  || |  |  | |  |||| |||  ||| |||     |
Db      20721 GTTATTTT-------TTTATTATTTTAATTTATTGTTTTTTTTTTTTTTATTTTTATAATAG 20773

Qy        361 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 420
              ||||  || |     || |   ||  || |||      |  ||   ||     ||||
Db      20774 TTTTTATTTTGGGTTTTAGTTTTTTTTATTTTTTTTATAGATTTTAAATTTTGATTTAGG 20833

Qy        421 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 480
              |  |||||| |  |  ||  | | |      |||||  ||       |    | || |
Db      20834 GGTTTGATTATAGTTATTTAAATGTT--TTTTAATGGTGTTTAAAATTTAGATTTTTTAA 20891

Qy        481 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 540
              |     |||| | |   || |||     |  | |  |        ||| |  ||| |||  |
Db      20892 GTTATTTTTTTTTTTAAAGTTATTATATTTTTTTTTTAATATTTTTATTTTATTTTGTGG 20951

Qy        541 AAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTAT 600
              |  |   |  |||  ||    |  |   | |||||| |||  ||||       ||
Db      20952 TTTGGAAGTTGTTTATTTAGTTATTTTTTTATATTTTTATTTGTTTTTATAGGTTTTTTG 21011

Qy        601 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAAT-AT 659
              ||   ||        ||| | | ||    ||| |||||||  |    ||        |||| ||
Db      21012 ATATTTTAGTTAGTTTTGATTTAAATTTTAATTTTTTTTTATTATATTATTTTTAATAAT 21071

Qy        660 TGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGG 719
              |    |       |  |||     ||||||||    || ||  |   |   | ||  | |
Db      21072 TTTTAGGGTTTATTTTTTATTACGTTTGGTTTTTTAGGTTATTTTATTATTTTTTAAAGT 21131

Qy        720 TATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCC 779
              |          |  ||||  |     |     |  |  |   ||||  |    ||
Db      21132 TTGTTTTTTAAATTATTTTTATTTTTTGATTTATATTTGTGGGGTATTATTTTAGTTTTT 21191

Qy        780 ATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGT 839
              ||  || |||  |  | |     ||  || || |||  | || |  || || |||  | |
Db      21192 ATTTTTTATTTTTTTTAATGGATATTTTTGTTGATTTAGTTTTTAAATTAGAATTTATAT 21251

Qy        840 GATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTT 899
              |    |  ||    | | |  | | ||  || |||||| || |||  | |   ||| |
Db      21252 TTTTTTGGTTTTTAGATTTTTATATTTGGTTATTATTTTTTATAAGAGGTTTTTGTTCGT 21311

Qy        900 ATTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGT 959
              |||| |      |  |  |  ||||| | | ||  || | ||  |  | |  |  | |
Db      21312 TTTATTTTTTTTATTTTGGTTAGTTTTTGTTTTTATTTTATTTGTATTTTTTTAGATAAT 21371

Qy        960 GTTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGT 1019
              |||  |  |              ||||     ||||  |    || ||       ||   ||
Db      21372 TTTATATATTTTTAGTTTTTTTTTTTTTATATTTTTTTTTGATGTTTAATTATTTTTTGT 21431

Qy       1020 TGGAGTTGTTTATGGATTTGCAAAAATT 1047
              | |||| |||  |  |||     |||||
Db      21432 AGAAGTTTTTTTAGTGTTTTTTTAAATT 21459


RESULT 12
US-10-311-455-254
; Sequence 254, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 254
;   LENGTH: 13511
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 8889
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-254

  Query Match             9.7%;  Score 101.8;  DB 6;  Length 13511;
  Best Local Similarity   47.3%;  Pred. No. 5.2e-08;
  Matches  481;  Conservative    0;  Mismatches  517;  Indels   19;  Gaps    5;

Qy         41 TTGTTTTTTCTTGCACTATTTACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATT 100
              || |||||| |      | |||  |||||||  | |   |||  |  |        | | ||
Db      10966 TTTTTTTTTATGATTTTTTTTTTATATTTATTTAGGATTTTATTTTTATTTTTTTTTTTT 11025

Qy        101 TAGATAAAAGTATTTTAATTCTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTC 160
              ||  | | | | | || ||| || |  | ||  |  || |||| ||| | |||    ||
Db      11026 TATTTGAGATTTTATTTATTTTTTATGGGGTTATTGATTATTTTTGTATTGATTGTTTTTT 11085

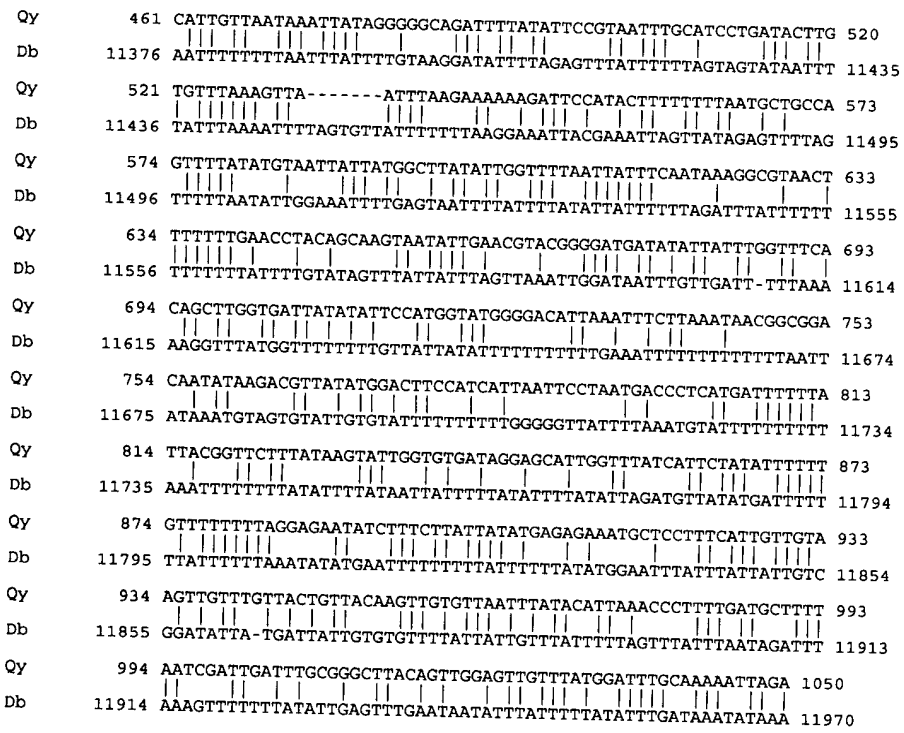
Qy        161 CTGAGCTTAATGTAAACGGTTTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGG 220
              ||  | | |     ||| |   |  | |  |  || | |  |    | |  | |
Db      11086 AGTTTATTTGTTTTAGTATTTTAATGTTTAGATAGATAATAGAAATATTATATTATAGAG 11145

Qy        221 TCTTTATCGCTTTTCAAAAACCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTT 280
              | |||   | ||     |     ||||  || | | |  | | | ||  ||  |
Db      11146 ATTATATTTTTATTTTTTGATTTTTTTTTATTTTTTTAGATGTTAAATATA-TGAGGTAAT 11204

Qy        281 TGAACTCTGCATTTAATTTTTTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTT 340
              |  | | |  ||||| ||||| | || ||  |  ||    || || |   | ||||  ||
Db      11205 TTTATTTTTTATTTAGTTTTTATTTTTGAAGTATTTTTTTTATGTATTTTTTATTTTTTT 11264

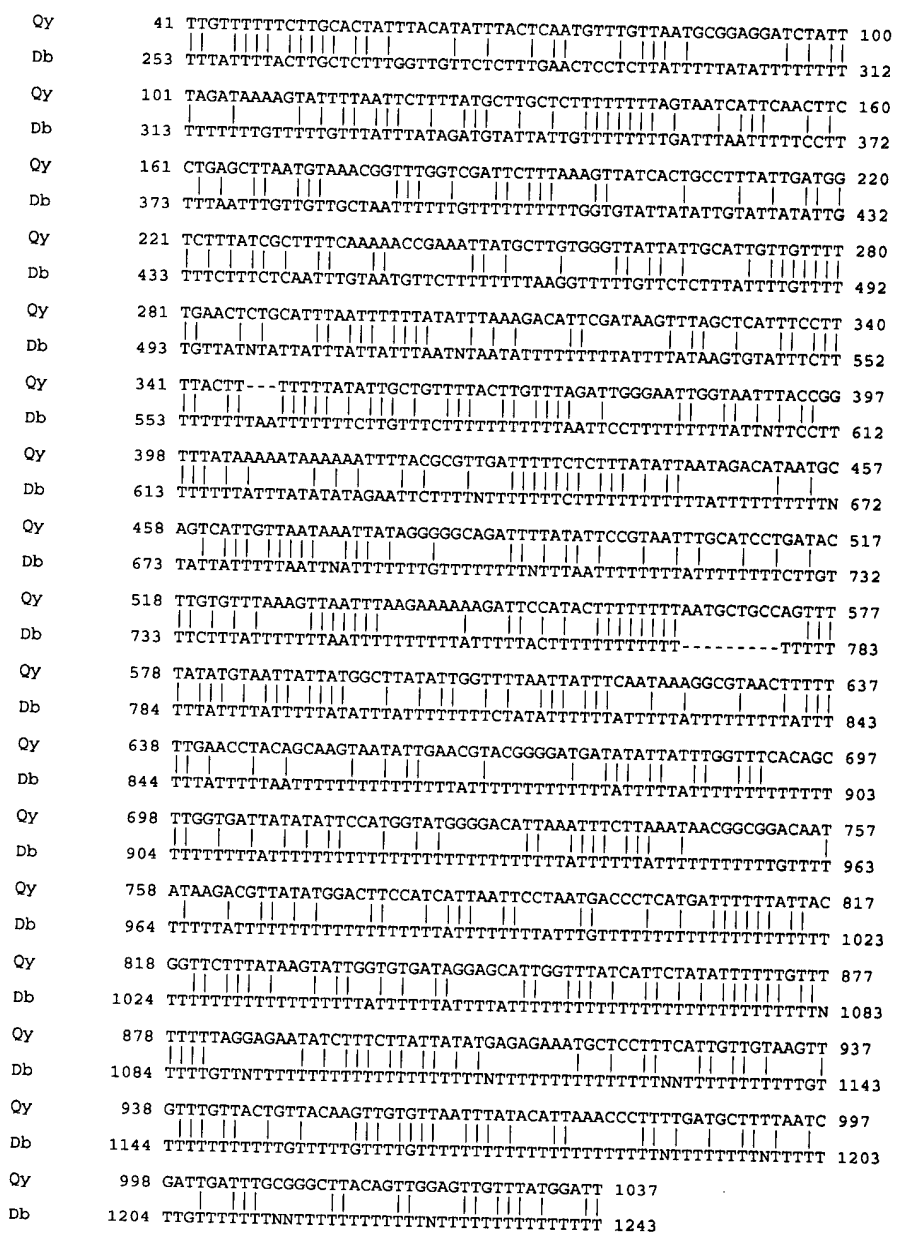
Qy        341 TTACTTTTTTTATATTGCTGTTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTT 400
              |  | || |||| ||  || ||     | ||     || |  |||    || |||    ||
Db      11265 ATTTTGTTATTATGTTTATGGTTATAGTTTTATTTTTAGTTATTATGAAATTATATTTTG 11324

Qy        401 ATAAAAATAAAAAATTTTACGCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGT 460
              | |  ||  |    ||        |  ||  |||  | |  ||||  |  |   ||
Db      11325 AAGATTGTATTAGTGGTT---------TAGTTGGTTTTTGAGTTTTAGAATTTTTTTTGT 11375
```



RESULT 13
US-10-425-115-172717
; Sequence 172717, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 172717
; LENGTH: 1243
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1243)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_89100C.1
US-10-425-115-172717

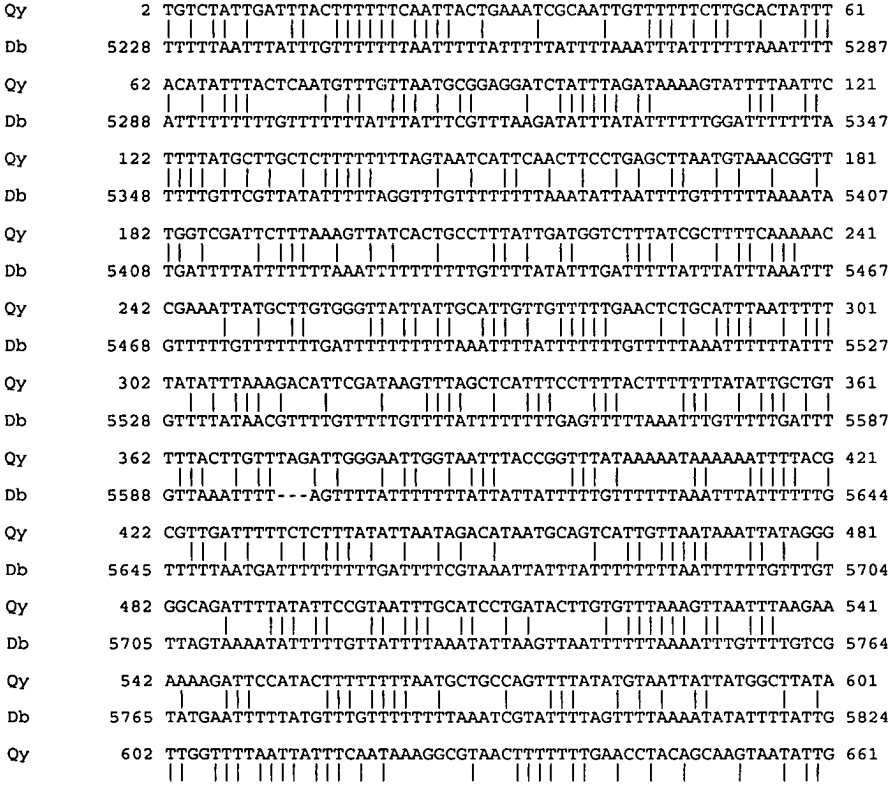
Query Match 9.6%; Score 100.6; DB 8; Length 1243;
Best Local Similarity 44.7%; Pred. No. 4e-08;
Matches 447; Conservative 0; Mismatches 541; Indels 12; Gaps 2;

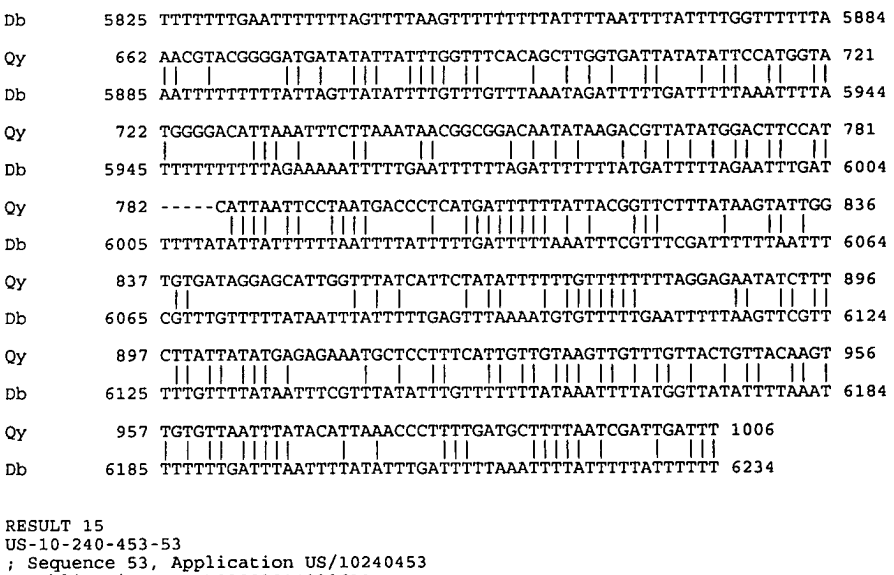


RESULT 14
US-10-239-676-51
; Sequence 51, Application US/10239676
; Publication No. US20030082609A1

; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 51
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-51

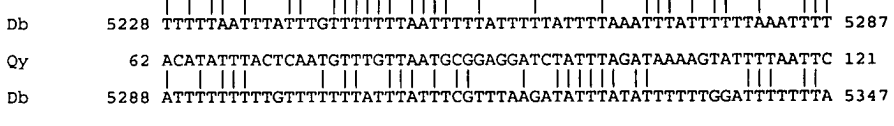
Query Match 9.5%; Score 100.4; DB 5; Length 9539;
Best Local Similarity 45.1%; Pred. No. 8.3e-08;
Matches 456; Conservative 0; Mismatches 546; Indels 8; Gaps 2;

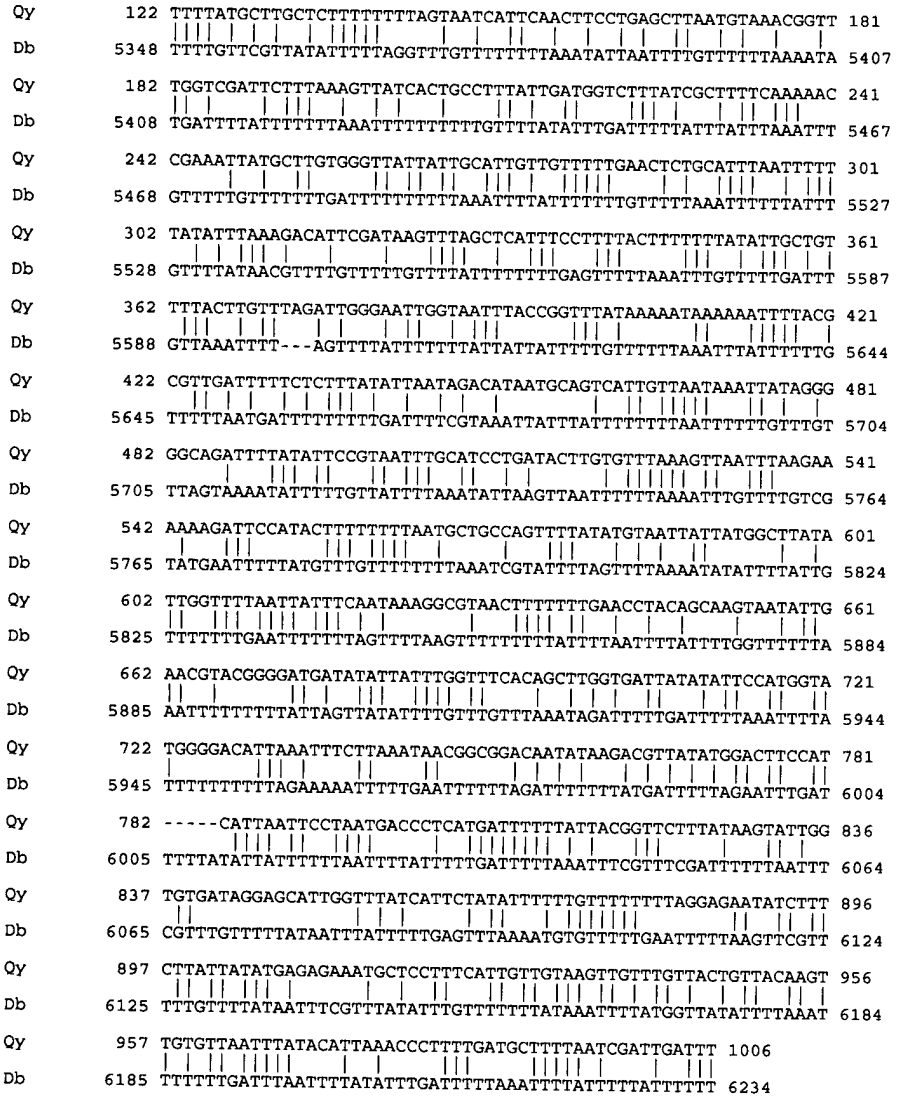




RESULT 15
US-10-240-453-53
; Sequence 53, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: With DNA TRanscription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 53
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-53

Query Match 9.5%; Score 100.4; DB 6; Length 9539;
Best Local Similarity 45.1%; Pred. No. 8.3e-08;
Matches 456; Conservative 0; Mismatches 546; Indels 8; Gaps 2;





Search completed: May 11, 2006, 07:09:27
Job time : 701.011 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:23:58 ; Search time 544.139 Seconds
(without alignments)
7892.621 Million cell updates/sec

Title: US-09-423-093B-1_COPY_9901_10953
Perfect score: 1053
Sequence: 1 atgtctattgatttactttt.....gatttgcaaaaattagataa 1053

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	81.4	7.7	8467	17	US-11-011-332A-96	Sequence 96, Appl
2	78	7.4	6156	10	US-10-240-708-60	Sequence 60, Appl
3	77.8	7.4	6583	10	US-10-240-708-26	Sequence 26, Appl
4	77.8	7.4	6669	10	US-10-240-708-6	Sequence 6, Appli
5	73.8	7.0	11049	10	US-10-240-708-23	Sequence 23, Appl
6	73.4	7.0	5455	10	US-10-240-708-33	Sequence 33, Appl
7	73.2	7.0	6306	10	US-10-240-708-50	Sequence 50, Appl
8	73.2	7.0	28536	17	US-11-011-332A-156	Sequence 156, App
9	72.4	6.9	10619	10	US-10-240-708-4	Sequence 4, Appli
10	72.4	6.9	28536	17	US-11-011-332A-155	Sequence 155, App
11	71.8	6.8	5501	10	US-10-240-708-38	Sequence 38, Appl
12	71.4	6.8	11049	10	US-10-240-708-22	Sequence 22, Appl
13	71	6.7	6499	17	US-11-011-332A-91	Sequence 91, Appl
14	71	6.7	28536	17	US-11-011-332A-151	Sequence 151, App
15	69.2	6.6	5535	10	US-10-240-708-17	Sequence 17, Appl
16	68.8	6.5	19233	10	US-10-240-708-45	Sequence 45, Appl
17	68.4	6.5	10467	10	US-10-240-708-2	Sequence 2, Appli
18	67.8	6.4	6070	10	US-10-240-708-10	Sequence 10, Appl
19	67.8	6.4	49979	10	US-10-995-561-13443	Sequence 13443, A
20	67.2	6.4	8467	17	US-11-011-332A-82	Sequence 82, Appl
21	67	6.4	5666	10	US-10-240-708-29	Sequence 29, Appl
22	66.8	6.3	5562	10	US-10-240-708-63	Sequence 63, Appl
23	66.2	6.3	6499	17	US-11-011-332A-77	Sequence 77, Appl
24	66	6.3	8537	10	US-10-240-708-41	Sequence 41, Appl
25	65.8	6.2	5610	10	US-10-240-708-53	Sequence 53, Appl
26	65.8	6.2	6040	10	US-10-240-708-69	Sequence 69, Appl
27	64.8	6.2	5152	10	US-10-240-708-73	Sequence 73, Appl
28	64.2	6.1	6866	10	US-10-240-708-20	Sequence 20, Appl
29	64	6.1	5610	10	US-10-240-708-54	Sequence 54, Appl
30	64	6.1	6326	10	US-10-240-708-57	Sequence 57, Appl
31	64	6.1	11049	10	US-10-240-708-21	Sequence 21, Appl
c 32	63.8	6.1	1230	7	US-09-925-065A-77681	Sequence 77681, A
c 33	63.8	6.1	1230	7	US-09-925-065A-77683	Sequence 77683, A
c 34	63.8	6.1	1230	11	US-10-301-480-178920	Sequence 178920,
c 35	63.8	6.1	1230	11	US-10-301-480-178922	Sequence 178922,
c 36	63.8	6.1	1230	12	US-10-301-480-792329	Sequence 792329,
c 37	63.8	6.1	1230	12	US-10-301-480-792331	Sequence 792331,
38	63.6	6.0	8961	10	US-10-240-708-80	Sequence 80, Appl
39	63.4	6.0	8093	10	US-10-240-708-31	Sequence 31, Appl
40	63.4	6.0	9001	17	US-11-011-332A-4	Sequence 4, Appli
41	63.2	6.0	9347	10	US-10-240-708-35	Sequence 35, Appl
42	62.8	6.0	8093	10	US-10-240-708-32	Sequence 32, Appl
43	62.8	6.0	171486	17	US-11-121-086-105	Sequence 105, App
c 44	62.6	5.9	1230	7	US-09-925-065A-77684	Sequence 77684, A
c 45	62.6	5.9	1230	11	US-10-301-480-178923	Sequence 178923,

ALIGNMENTS

RESULT 1
US-11-011-332A-96
; Sequence 96, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
; APPLICANT: Foekens, John
; APPLICANT: Harbeck, Nadia
; APPLICANT: Koenig, Thomas
; APPLICANT: Maier, Sabine
; APPLICANT: Martens, John
; APPLICANT: Model, Fabian
; APPLICANT: Nimmrich, Inko
; APPLICANT: Rujan, Tamas
; APPLICANT: Schmitt, Manfred
; APPLICANT: Lesche, Ralf
; APPLICANT: Dietrich, Dimo
; APPLICANT: Mueller, Volkmar
; APPLICANT: Kluth, Antje
; APPLICANT: Schwope, Ina
; APPLICANT: Hartmann, Oliver
; APPLICANT: Adorjan, Peter
; TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SUR
; TITLE OF INVENTION: BREAST CELL PROLIFERATIVE DISORDER PATIENTS
; FILE REFERENCE: 47675-99
; CURRENT APPLICATION NUMBER: US/11/011,332A
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 10/517,741
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: PCT/EP2003/010881
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: DE 10245779.4
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: DE 10300096.8
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: DE 10317955.0
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: PCT/EP2004/014170

```
;  PRIOR FILING DATE: 2004-12-13
;  PRIOR APPLICATION NUMBER: EP 03090432.0
;  PRIOR FILING DATE: 2003-12-11
;  PRIOR APPLICATION NUMBER: EP 04090041.7
;  PRIOR FILING DATE: 2004-02-10
;  PRIOR APPLICATION NUMBER: EP 04090380.9
;  PRIOR FILING DATE: 2004-09-30
;  PRIOR APPLICATION NUMBER: EP 04090127.4
;  PRIOR FILING DATE: 2004-04-01
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 96
;   LENGTH: 8467
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-96

  Query Match             7.7%;  Score 81.4;  DB 17;  Length 8467;
  Best Local Similarity   48.3%;  Pred. No. 0.0044;
  Matches  332;  Conservative    0;  Mismatches  336;  Indels   19;  Gaps    3;

Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
Db       3258 ATTTTTTTTGATTTTTTTTTTTTTTTTTTTATATATGTTTTTTTTTTTTTTGTTTTTATT 3317

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
Db       3318 TTTTTTTTATTTTTTTTTTTTTTTTTGTTTTTTTTTTATTTTTTTTTTTTTTTTTTTT 3377

Qy        121 CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 180
Db       3378 TTTTTTTTTTATTTTTTTTTTGTTTTTGAGTTTTTATTGATTGATTTTTTTTTATTTTAT 3437

Qy        181 TTGGTCGATTCTTTAAA-GTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAA 239
Db       3438 TTGTTTTTTTTTTAATGTGTTAATTTTTGTTTTATTTTTGATTTTTTTAGGTATTGGGAG 3497

Qy        240 ACCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTA---- 295
Db       3498 GTGGGATGGGGGTGTGTGTTTTTTTTTAGGAGTTTTGTTTTTTTTAAGATTTATAGAAATT 3557

Qy        296 ---ATTTTTTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTA 352
Db       3558 AGGATTTGTTTTTATTTAAAATTTTATGTATTTTAAGTTTTTTTTAGATAATATATTTTA 3617

Qy        353 TATTGCTGTTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAA 412
Db       3618 ATTTTTTGGGTTGATTAGTTTTTTTGTGTAGAGGTAGTTGAGAGGTTT----------T 3666

Qy        413 AATTTTACGCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAA 472
Db       3667 GTTTTGTAGAGGGAAAAGAGTTTTTTATTTTTTTATTTATTATATAGGTAAATTTATTTG 3726

Qy        473 ATTATAGGGGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTA 532
Db       3727 GTTATTGGTTGAAGGTATAGTTTTGTTTTTGTGGGGAATTGGTGGTTAGGATATAATAGT 3786

Qy        533 ATTTAAGAAAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTA 592
Db       3787 GTTTTTGGAGTTTATTTTTGGTTTTGGTGTTGGTGTAGGGATTTTTTGATTGGGTTTGAG 3846

Qy        593 TGGCTTATATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAA 652
Db       3847 GGGTTTGGGTTAGTTTTAATGTTATTATTTATAGTGAGGGTAGGGTGTAAGGTTGAGAAG 3906

Qy        653 GTAATATTGAACGTACGGGGATGATAT 679
Db       3907 GTTATATTTATTGTTTTGGGAGGATGT 3933

RESULT 2
```

```
US-10-240-708-60
; Sequence 60, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/240,708
;  CURRENT FILING DATE:  2002-10-03
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 60
;   LENGTH: 6156
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-60

  Query Match             7.4%;  Score 78;  DB 10;  Length 6156;
  Best Local Similarity   44.2%;  Pred. No. 0.013;
  Matches  411;  Conservative    0;  Mismatches  515;  Indels    4;  Gaps    2;

Qy         79 TTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTTATGCTTGCTCTTT 138
Db       3757 TTATTAAAGGATTAAATTTTATTATTTAATTAAATTTTAAAAATATAATGTTATAATATT 3816

Qy        139 TTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGTCGATTCTTTAAAG 198
Db       3817 TTTTAAGAATATTATAGTATTTTATAATTTTAAATTTTGTGGTAATTTTATATTTTATAG 3876

Qy        199 TTAT--CACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAAATTATGCTTGT 256
Db       3877 TTGTTATATTAAAGTTATGGGAGAGTTTTAATTCGTTATTTTAAGTTTAGTTATAATTTT 3936

Qy        257 GGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTAAAGACAT 316
Db       3937 TTTTTTATTTTTATTTAATGTTTTATATATTTTTATTGATTTTTTAAAGTTTTAGGAATT 3996

Qy        317 TCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACTTGTTTAGAT 376
Db       3997 ATTTTTATGTTATATTGAGTGGGAGGTTTTATTTAATGGTTTTGTTATGGAAATTTTGGT 4056

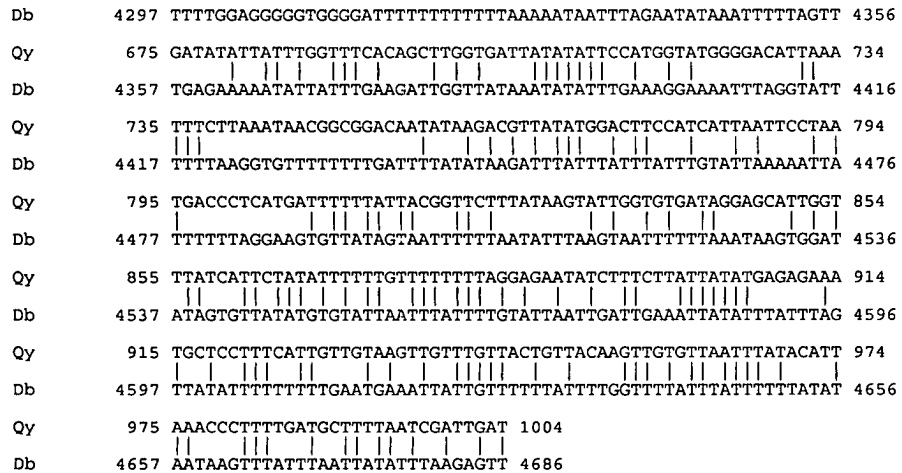
Qy        377 TGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGATTTTTCTCT 436
Db       4057 GGTGTAGTTTTTTATGTTAATTTAAAGATTGTAAGAGGTTTTTTTTTTAAAATATATTAT 4116

Qy        437 TTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCAGATTT--TATA 494
Db       4117 GGATTTAAGGTGTTATATTTTATGTTTTTTTTTTAATATTTTTTTAGAAGGGTTGGTGTA 4176

Qy        495 TTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGATTCCATA 554
Db       4177 TTAATTAATTGATATAAAAATATATTTGTAATAATATATTTTAGTAGTATTTTTGTAATT 4236

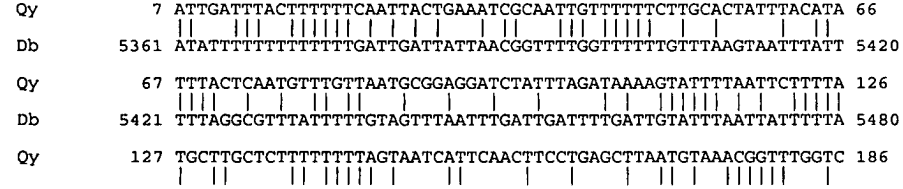
Qy        555 CTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGGTTTTAATTA 614
Db       4237 GTTATTTTTAAAATTATGTTAATATTTTTAATATAATTTTGTATTTTATAGGTGTTTTAA 4296

Qy        615 TTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTACGGGGAT 674
```



RESULT 3
US-10-240-708-26
; Sequence 26, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 26
; LENGTH: 6583
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-26

Query Match 7.4%; Score 77.8; DB 10; Length 6583;
Best Local Similarity 43.8%; Pred. No. 0.014;
Matches 445; Conservative 0; Mismatches 562; Indels 10; Gaps 2;





RESULT 4
US-10-240-708-6
; Sequence 6, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971

; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 6
; LENGTH: 6669
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-6

Query Match 7.4%; Score 77.8; DB 10; Length 6669;
Best Local Similarity 43.2%; Pred. No. 0.014;
Matches 422; Conservative 0; Mismatches 552; Indels 3; Gaps 1;

```
Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
Db       3905 ATGGTTTTTTAGATAGTTGTTTTCGTAATTTATTTTATTTATTGTTGTTATAAATTTTTT 3964

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
Db       3965 TGAAGTTTTTATATTATATTTTTTAGTTTGAAAGTTGAGATAATTTTATTTTTATTAAAA 4024

Qy        121 CTTTTATGCTTGCTCTTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 180
Db       4025 ATTTTAGATGGGTAGGAGTATATAATATAAAAAAGTTTTTTTATGTTTAATGTTTTTATT 4084

Qy        181 TTGGTCG---ATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAA 237
Db       4085 TTTGTCGTTTATTATTTTATATTTTAAATAGATATATATCGTGTATATTTTTTTTATATG 4144

Qy        238 AAACCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAAT 297
Db       4145 AGGTTTATTATTTATTTATGATTTTTTTTTTTATTAGTATTTTTTCGTTATTTTTAGTTTT 4204

Qy        298 TTTTTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTG 357
Db       4205 TTTTGTTATTGGTTTTTTTATAATGTTTTTTTTTTTAGTAAATTTTTGGGTGTTTAGTATG 4264

Qy        358 CTGTTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTT 417
Db       4265 TTTTAAATATTGTGTTAATTGTTGAATATAGAGAAGTTAAGTTTTTTTTTTTTAGTGTTT 4324

Qy        418 TACGCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTAT 477
Db       4325 AGATAATTAGGTTTTTTTTAAATTTTAAATATATATATAAAAAAATATTTTTTTGGATT 4384

Qy        478 AGGGGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTA 537
Db       4385 TTATTAGATATTAGATATTTATTAAATAGAATATTTTGTTTTATGTTTTGTGTTATTTTT 4444

Qy        538 AGAAAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCT 597
Db       4445 TTATAAGTTAAATTTATTTAAAGAATAATGAATTTTATTTAGTTGTTGTTTTTATTTAT 4504

Qy        598 TATATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAAT 657
Db       4505 TTTTTATTTTTTAGGATATTGTATTTTGATTTTTGTATTTTATTATTTTATTTAAGTAGT 4564

Qy        658 ATTGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCAT 717
Db       4565 TTTTATTGTTTTGTTGGGATTTTATAAAATGTTGTGTGGGTTAGAATATATTAAAGTGTT 4624

Qy        718 GGTATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTT 777
Db       4625 TGTTGTTTTTGGGTTATTAGTTTAGTTTTTATTTATTTATTTAAGGTGTTTTGTGGAATT 4684
```

```
Qy        778 CCATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGT 837
Db       4685 GATTTTATTTATTTTTTTAATTTGATTTTAATTTTTTAATATTTGTTTTATGTTGTAGAT 4744

Qy        838 GTGATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTC 897
Db       4745 ATGTTAAACGAATTGAAGTTTTTTTGAAAATTTTTGTATGTATTTTATTTTTTTTTGGTT 4804

Qy        898 TTATTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTT 957
Db       4805 TTTTAGTATATTTTTATTTAAGGTTTAGGTTATGTTTGATTTTTGAAATTTTATTTAGTA 4864

Qy        958 GTGTTAATTTATACATT 974
Db       4865 AGTATAGTTAGTATTTT 4881
```

RESULT 5
US-10-240-708-23
; Sequence 23, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 23
; LENGTH: 11049
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-23

Query Match 7.0%; Score 73.8; DB 10; Length 11049;
Best Local Similarity 45.3%; Pred. No. 0.057;
Matches 349; Conservative 0; Mismatches 417; Indels 5; Gaps 2;

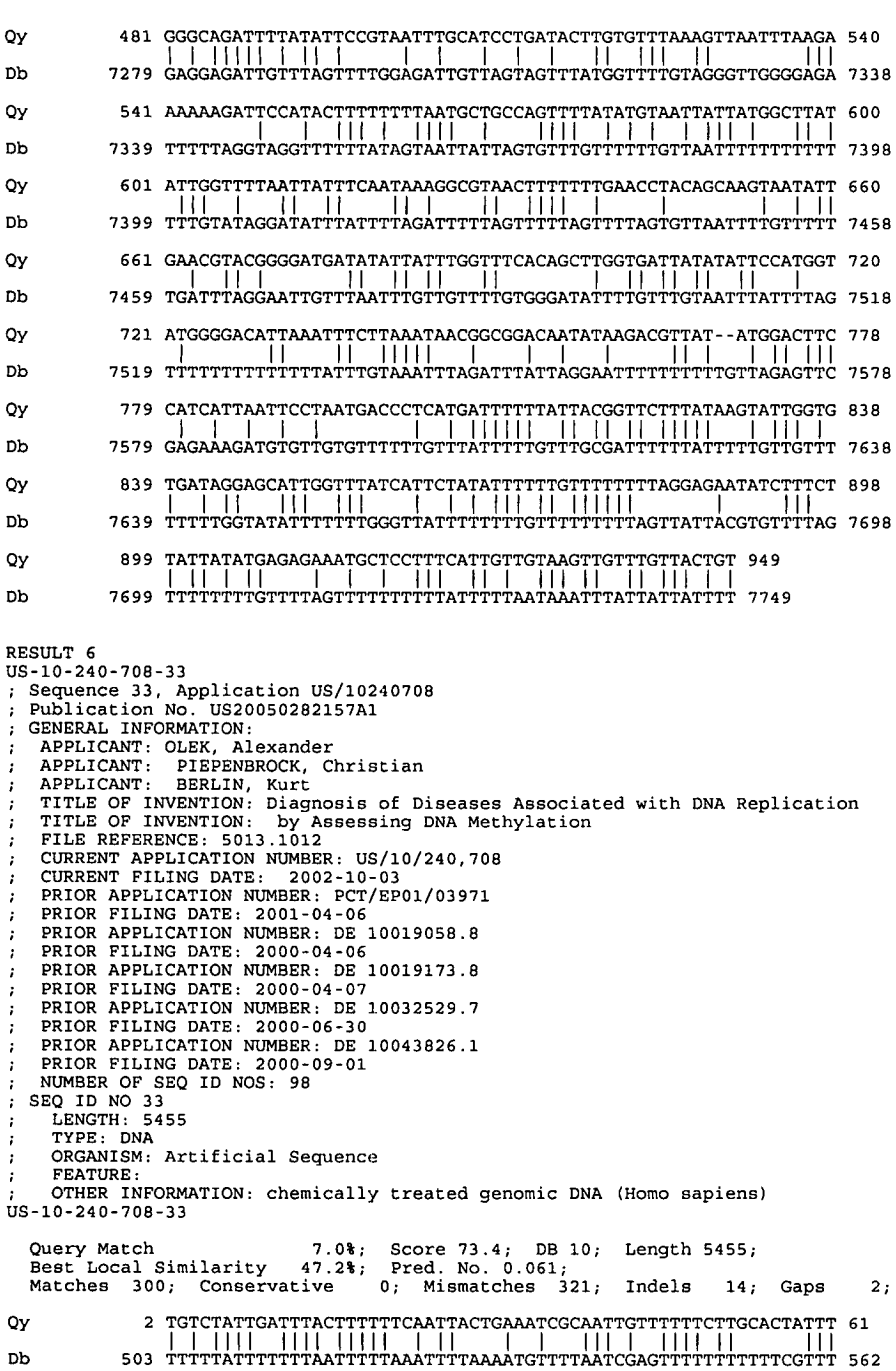
```
Qy        181 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 240
Db       6982 TTTTTGAAATTTTTTAATTTGATAAGAGTTTTTTGGTTTTTAATTTTTATTTTTAGAAGA 7041

Qy        241 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 300
Db       7042 TTTGGTTTTTTGGTTATTGTTAGATATTAAGTGATTTTTTGAATTTTTTTTTTTTATTTTT 7101

Qy        301 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTTATATTGCTG 360
Db       7102 AGGTTTTTTTTGTTTTTTTGTTTGTTTGTTTTTGTTTTTTTAAATTTATTTAGTTTGGAT 7161

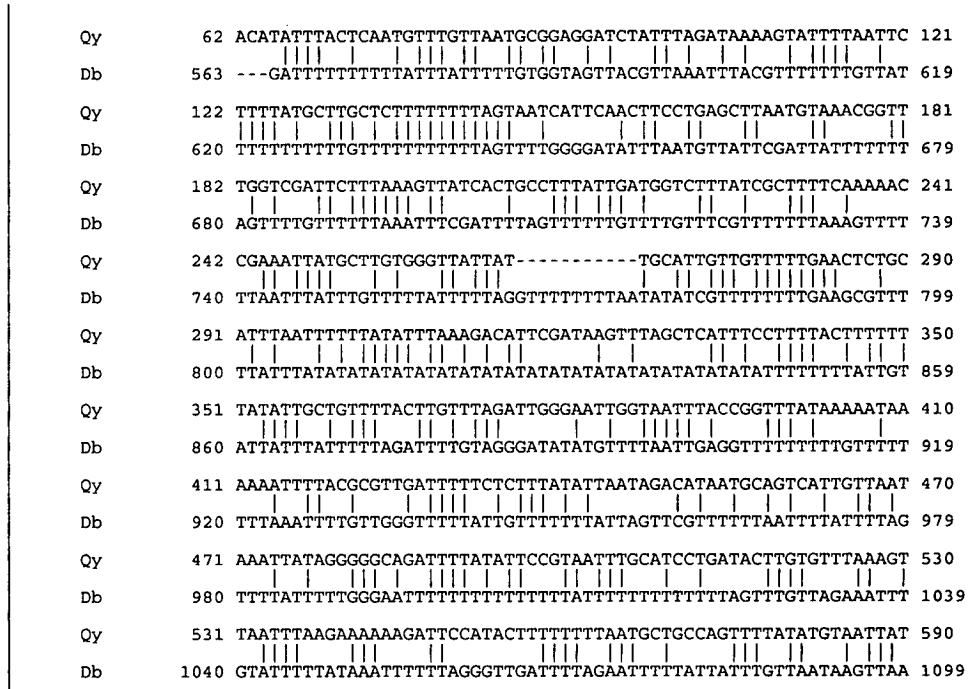
Qy        361 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 420
Db       7162 TTTGAAGTTAGTTGTTTTAG---TTATTTTTTTTAGTATTTTTAATTTCGTGATTTTTTT 7218

Qy        421 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 480
Db       7219 GCGTTATTATTTTTGTTTTGAAAATTTTTAATAGTGATATTTTTTTTTAGGATTATTGAAG 7278
```



RESULT 6
US-10-240-708-33
; Sequence 33, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 33
; LENGTH: 5455
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-33

Query Match 7.0%; Score 73.4; DB 10; Length 5455;
Best Local Similarity 47.2%; Pred. No. 0.061;
Matches 300; Conservative 0; Mismatches 321; Indels 14; Gaps 2;



RESULT 7
US-10-240-708-50
; Sequence 50, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 50
; LENGTH: 6306
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-50

```
Query Match             7.0%;  Score 73.2;  DB 10;  Length 6306;
Best Local Similarity   45.0%;  Pred. No. 0.066;
Matches  447;  Conservative    0;  Mismatches  533;  Indels   14;  Gaps    4;

Qy          2 TGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTT 61
Db       5194 TTTTTTTTTATTTTATTTTTTTTTTTTTTTTTATTAGGAA--GATTTTTTTAATATAGTT 5251

Qy         62 ACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTC 121
Db       5252 AAATATTTTTTTTTTTAAGTTTAATGTAGGAGAAAAGTTTATTAAAATTTTATTTATGAA 5311

Qy        122 TTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTT 181
Db       5312 TTTTTTGATATATCGTATTTTTAAAGGTATTTTATTTTTTTTAAAATGTTTAGAGTTTTG 5371

Qy        182 TGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAAC 241
Db       5372 TTTTTTATATTTAATAGTTCGTTATTTAGAATTTATTTTTTATGTATGTTGTTAGATAAG 5431

Qy        242 CGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTT 301
Db       5432 TTTTTAATTTTTGTTTTGTTTTAATTTTAGTTGTTTTAATGTGATTTGATTGTATTATTT 5491

Qy        302 TATATTTAAAGACATTCGATAAGTTTAGCT----CATTTCCTTTTACTTTTTTTTATATTG 357
Db       5492 TTTTGTTATTTTTATTTGTTATTTTAAGTTTATAGATATGTGTGGAGTTAATTTTATTTT 5551

Qy        358 CTGTTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTT 417
Db       5552 ATGGATTTATTTTTTTATTTTTGAATTATTATTTATTTTTTTTAATAGTGTAGTTTATTA 5611

Qy        418 TACGCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAAT------A 471
Db       5612 TAATTTTTGATGTTTGGGGTTGTGTTTTCGGGAATGTTTTGGTTATTTTTGTTTTTTTGT 5671

Qy        472 AATTATAGGGGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTT 531
Db       5672 TTTTTTGTATAGAATGTAGTTTATGAATTTTTAGGAAAAATTGTGTTGGAATTTTGATTA 5731

Qy        532 AATTTAAGAAAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATT 591
Db       5732 AAATTGTATTGTATTTTTGGATGTATTTTAGGAAAAGTAATATTTGAATGTTAATTTTTA 5791

Qy        592 ATGGCTTATATTGGTTTTAATTATTTCAATAAAGGCGT--AACTTTTTTTGAACCTACAG 649
Db       5792 TTATTTTTGAGTTATTTTGGTTTTTTTAAAATTTTTTTAAAAGGTTTTATATTTTTTTAT 5851

Qy        650 CAAGTAATATTGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATA 709
Db       5852 ATAAAAGTTTTGATTTTTAGTTTAGATTTATTTTTAGATATTTTGTTTTTTGTTGTTATTA 5911

Qy        710 TATTCCATGGTATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTAT 769
Db       5912 TGAATGGTATTATTTTTTAAATTAAAGATTTTAATTGATTGTTTTGGTATATAGAAATGT 5971

Qy        770 ATGGACTTCCATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAA 829
Db       5972 TATTTTAAGAATTTTTTATTAGTTTTTTTTGGAATTTTTTATTTATATTTGAATAATTAT 6031

Qy        830 GTATTGGTGTGATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAA 889
Db       6032 AGTTTTGTTTTTTAAGAATTTTTATATATTTTATTTATTTTTTTTTATTTTTTGTTTAGGA 6091

Qy        890 TATCTTTCTTATTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGT 949
Db       6092 TAGAGTTTTTAGTATTATTTTGAATAGAAATAGTGTTTAAGTTAAGATGGGGTTTAGTTG 6151

Qy        950 TACAAGTTGTGTTAATTTATACATTAAACCCTTT 983
Db       6152 TATAAATAAAAGTATTTTTAATGTATAAACGTTT 6185
```

```
RESULT 8
US-11-011-332A-156
; Sequence 156, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
;  APPLICANT: Foekens, John
;  APPLICANT:  Harbeck, Nadia
;  APPLICANT:  Koenig, Thomas
;  APPLICANT:  Maier, Sabine
;  APPLICANT:  Martens, John
;  APPLICANT:  Model, Fabian
;  APPLICANT:  Nimmrich, Inko
;  APPLICANT:  Rujan, Tamas
;  APPLICANT:  Schmitt, Manfred
;  APPLICANT:  Lesche, Ralf
;  APPLICANT:  Dietrich, Dimo
;  APPLICANT:  Mueller, Volkmar
;  APPLICANT:  Kluth, Antje
;  APPLICANT:  Schwope, Ina
;  APPLICANT:  Hartmann, Oliver
;  APPLICANT:  Adorjan, Peter
;  TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SUF
;  TITLE OF INVENTION:  BREAST CELL PROLIFERATIVE DISORDER PATIENTS
;  FILE REFERENCE: 47675-99
;  CURRENT APPLICATION NUMBER: US/11/011,332A
;  CURRENT FILING DATE:  2004-12-13
;  PRIOR APPLICATION NUMBER: US 10/517,741
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: PCT/EP2003/010881
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: DE 10245779.4
;  PRIOR FILING DATE: 2002-10-01
;  PRIOR APPLICATION NUMBER: DE 10300096.8
;  PRIOR FILING DATE: 2003-01-07
;  PRIOR APPLICATION NUMBER: DE 10317955.0
;  PRIOR FILING DATE: 2003-04-17
;  PRIOR APPLICATION NUMBER: PCT/EP2004/014170
;  PRIOR FILING DATE: 2004-12-13
;  PRIOR APPLICATION NUMBER: EP 03090432.0
;  PRIOR FILING DATE: 2003-12-11
;  PRIOR APPLICATION NUMBER: EP 04090041.7
;  PRIOR FILING DATE: 2004-02-10
;  PRIOR APPLICATION NUMBER: EP 04090380.9
;  PRIOR FILING DATE: 2004-09-30
;  PRIOR APPLICATION NUMBER: EP 04090127.4
;  PRIOR FILING DATE: 2004-04-01
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 156
;   LENGTH: 28536
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-156

Query Match             7.0%;  Score 73.2;  DB 17;  Length 28536;
Best Local Similarity   45.1%;  Pred. No. 0.076;
Matches  397;  Conservative    0;  Mismatches  473;  Indels   10;  Gaps    3;

Qy         66 ATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTT 125
Db       5937 ATTTGTTGAATGTAGGTATAGTAGAGGAAATTTTAGATTTTATGAGTGTTTGAGTTTTGT 5996

Qy        126 ATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGT 185
Db       5997 TTAGTGTAAATTTTTTGTGAATATTGGGTTAGTGTGTGGTTGTGTTTATTTGTGTGTTGA 6056

Qy        186 CGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCA--AAAACCG 243
Db       6057 TATTTTTAGTATGTTTGGTTTATTTGTTTTGATTTTGGGTGTGGTGTTTTAGTTAAGTTG 6116
```

```
Qy        244 AAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTA 303
Db       6117 GGTTTAGTGTTTTGGTTTTTTTTAGTTGATAAGTTTAGTTTGTTTGTTTTGGTTGTGGT 6176

Qy        304 TATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTTATATTGCTGTTT 363
Db       6177 TTTTTTATTTTTTTTTTATTAGTTTATTTTATTTTTTTAGATTTTTTTTTATTTATTTTTT 6236

Qy        364 TACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCG 423
Db       6237 TTTATTTTTATTGTGTTTATTTTTATTTTTGTTTTTTATTGGTTTTTTATTTTTTTTTTT 6296

Qy        424 TTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGG 483
Db       6297 TTGTAGTTTTTTTTTGTTGTGATTTTTTTTT----TAATTTTGTAGGTTTGAAAGAAGG 6352

Qy        484 CAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAA 543
Db       6353 TTATATATGTATGTTTATATTTATATTTTATATGTTTTGTTTTAAATAATTTTATGAATA 6412

Qy        544 AAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATT 603
Db       6413 TTGTTTTTTGTTTTGTTTTTTGGGTT----ATTTTTTTTGTTGTTTTTTTTTAGTTTGTT 6468

Qy        604 GGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAA 663
Db       6469 TTGATTTGTTTTTTAAAAGTATGTTTTTGTTTTTTTGTTGTTTTGGTGTTTTTTTTTTGA 6528

Qy        664 CGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATG 723
Db       6529 TTTATTAGGGTTGTTGGGTTGGTGTAGATTGTTTTTTTTTTTTTTTTTATTTTATTTTTTT 6588

Qy        724 GGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCA 783
Db       6589 TTTTGGTTTTTTTTTTTATAGTGGGAGTTTGTGTTTTTGTTTTTTGGTTGGTTTTTAAGT 6648

Qy        784 TTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATA 843
Db       6649 GTTTTGTTAGGTTTTTTTTTTTTTTGTTTTTTTGGTTTTGGTTTTTGATTTTTTGGTTTG 6708

Qy        844 GGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTATTA 903
Db       6709 TTGGTATTTGTTTTTTTTTTTTTGTTTTGTTTTTTGTTGTTTTTGTTTGTTTTTTTTGGTG 6768

Qy        904 TATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGT 943
Db       6769 TTTGTTTGGGTGTTGTGTTTGTTTTTGGATTGTTAGTTGT 6808
```

```
RESULT 9
US-10-240-708-4
; Sequence 4, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/240,708
;  CURRENT FILING DATE:  2002-10-03
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
```

```
; SEQ ID NO 4
;   LENGTH: 10619
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-4

  Query Match             6.9%;  Score 72.4;  DB 10;  Length 10619;
  Best Local Similarity   44.5%;  Pred. No. 0.091;
  Matches  462;  Conservative    0;  Mismatches  566;  Indels   10;  Gaps    4;

Qy          1 ATGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATT 60
Db       8367 AAGGTTAAGAATTTTTGTTTTAGAATTTAAGGAAAGTTAATTATTATTTATTG--TTTTT 8424

Qy         61 TACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 120
Db       8425 TTAATATTTTAGTTTGGATGGTTTATATGTATTTTTATTTATGTTTAAAATTGTTTTAGT 8484

Qy        121 CTTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 180
Db       8485 GTTTTTTATTTTAGGAAATGGTTTATTATTTTTTGTATTATTTAGGT-----TAAATGTT 8539

Qy        181 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 240
Db       8540 TTGTAGTTAATTTGATTTTTTTTTTTTGTTTTATGTTTAATTTATTTTCGAATTTTAATT 8599

Qy        241 CCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTT 300
Db       8600 TTTTGATTTTGATTTTTTTTTATGGTTTAGTTGTTATTTTTTTTGGTGGTTTTGTATTTT 8659

Qy        301 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTG 360
Db       8660 TTATTAGAATGTAAGTTTTATGAGGGTAGGAGTTTGTGTTTGTTTTGTTGTAATAGTTTT 8719

Qy        361 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 420
Db       8720 AGTGTTTGGAATAGTGTTAGGTATTTATTAATATTTGTTGAATGAATGATTGAATTTTAT 8779

Qy        421 GCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 480
Db       8780 AGTTGGTTTTTTTTTCGTATAGGTTAATAGTTTAATTAAT-AATGTTATTTTATATTTGT 8838

Qy        481 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 540
Db       8839 ATATAATTAATTTATATTGTGTTTTAAATTTGAAATTTTTATGTTGATTTATAATTTTTA 8898

Qy        541 AAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATG--TAATTATTATGGCTT 598
Db       8899 TAGTATATTAGGTAAAGTTGGTGGTTATATAAAATATTATTAGGATAGGGTTGTTTGAAG 8958

Qy        599 ATATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATA 658
Db       8959 AATTATTTGATAATTTTTGTTAATAATATAAAAATAGAAAAAGAACGTTATTATGTAGTT 9018

Qy        659 TTGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATG 718
Db       9019 TATAAATAAATTTTAGGTGGTTTTATATTTTAATTTTTTTTTTTAAAAGGAATGAGTTGT 9078

Qy        719 GTATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTC 778
Db       9079 GTTTAGGTAGGTTAAATGTTTTATAGATAAGAGAAAAAAAAAAAAAAGTTGTTATAGAATT 9138

Qy        779 CATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTG 838
Db       9139 ATTTTATTTGTTATTATTATCGATTTTATTTTTTATTTTTTTTTATTTTTTTTGTTTTTTT 9198

Qy        839 TGATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCT 898
Db       9199 TCGTTTTGTTTTTTTTTTTTGTTTTTTTCGGTTTTGATAGTATTATTTTTGTTATATTAG 9258

Qy        899 TATTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTG 958
```

```
Db       9259 TTTTTTTTTGGTTGGATATGTAGTATTTTGTTTTTGGTATTTTTTTTTTAGTTGAGAAGT 9318

Qy        959 TGTTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAG 1018

Db       9319 TTTTTTTTTGGGTTGTTTGTTATTTGTAGTATGGTTATTTATGTTTTTTTTAGGTTTTTG 9378

Qy       1019 TTGGAGTTGTTTATGGAT 1036

Db       9379 GTATTATTGTTAGTGGAT 9396
```

RESULT 10
US-11-011-332A-155
; Sequence 155, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
; APPLICANT: Foekens, John
; APPLICANT: Harbeck, Nadia
; APPLICANT: Koenig, Thomas
; APPLICANT: Maier, Sabine
; APPLICANT: Martens, John
; APPLICANT: Model, Fabian
; APPLICANT: Nimmrich, Inko
; APPLICANT: Rujan, Tamas
; APPLICANT: Schmitt, Manfred
; APPLICANT: Lesche, Ralf
; APPLICANT: Dietrich, Dimo
; APPLICANT: Mueller, Volkmar
; APPLICANT: Kluth, Antje
; APPLICANT: Schwope, Ina
; APPLICANT: Hartmann, Oliver
; APPLICANT: Adorjan, Peter
; TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SU
; TITLE OF INVENTION: BREAST CELL PROLIFERATIVE DISORDER PATIENTS
; FILE REFERENCE: 47675-99
; CURRENT APPLICATION NUMBER: US/11/011,332A
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 10/517,741
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: PCT/EP2003/010881
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: DE 10245779.4
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: DE 10300096.8
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: DE 10317955.0
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: PCT/EP2004/014170
; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: EP 03090432.0
; PRIOR FILING DATE: 2003-12-11
; PRIOR APPLICATION NUMBER: EP 04090041.7
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: EP 04090380.9
; PRIOR FILING DATE: 2004-09-30
; PRIOR APPLICATION NUMBER: EP 04090127.4
; PRIOR FILING DATE: 2004-04-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 155
; LENGTH: 28536
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-155

Query Match 6.9%; Score 72.4; DB 17; Length 28536;
Best Local Similarity 44.6%; Pred. No. 0.099;
Matches 466; Conservative 0; Mismatches 566; Indels 14; Gaps 4;

```
Qy          2 TGTCTATTGATTTACTTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTT 61

Db      26427 TATTTTTTATTTGATTTTTTTTTTTTTGTTTATTTTTAATTTTTTAATTATTTAGAAATT 26486

Qy         62 ACATATTTACTCAATGTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTC 121

Db      26487 TTTTTATTTTTTAGTGGTTTTTTTTTTGTAGTAGTTTTTTATTTGAATTTTTTTTTTGTT 26546

Qy        122 TTTTATGCTTGCTCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTT----AATGTAAAC 177

Db      26547 TTTTTGTGGTAGGGTTTGTATATTGATTTTTTTGATTTTTGGTATATTTGGGTTTTTTGA 26606

Qy        178 GGTTTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAA 237

Db      26607 AATTTTTTAATTTTTTTAGATTTGAGGATGGTAGGTTTTATTTTTTTTTATTGTGTGTATA 26666

Qy        238 AAACCGAAATTATGCTTGTGGGTTATTATTGCATTGTTGTTTTTTGAACTCTGCATTTAAT 297

Db      26667 TATTTAGAGATATGAAAATTTATATAGATTGTTTTTAAATTTAGGGTATTTAATAGATGT 26726

Qy        298 TTTTTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTG 357

Db      26727 TTTTTTTTTAGTTTGTTTTTTGATTTGAAATGTTTGTTTGATTTTAATTTGGATATTATT 26786

Qy        358 CTGTTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTT 417

Db      26787 TTTTTTTGTTTTTTTTTTTTTAAAGTAGTTTGGATATGTGTGTAAGTGAGTTTAGAATAG 26846

Qy        418 TACGCGTTGATTTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTAT 477

Db      26847 TTTTATTTATATTTTTTATTAAATTGTAAATAAAAGAAGAATTAATGAAGTAGATTGGTA 26906

Qy        478 AGGGGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTA 537

Db      26907 TATAGATTGTATTAAGAGTTTGAATTTTTAGTTTTTGGATTTTTTATTTAATTTTGGTTG 26966

Qy        538 AGAAAAAAGATTCCATACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCT 597

Db      26967 TTATTTATATTGATAGAGTTATTTTAAGT----AGAGGTTTAGAGAAATTTGTATTGTGG 27022

Qy        598 TATATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAAT 657

Db      27023 GATAATAGGTAAAGTTATAGTAAAAAGTGGAATAATTTTAAAGTTATTTTATTAGAATGT 27082

Qy        658 ATTGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCAT 717

Db      27083 AAATTGTATTTTTGGGTTTTGTTTGTAATTATTTAGTTTTAATATATATAGAGTTAGATA 27142

Qy        718 GGTATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTT 777

Db      27143 GGAAAAAATAGGTTAATATAGTT-ATTGGTATTAGAGAAGATAAATTTTATGGGTTTTTT 27201

Qy        778 CCATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGT 837

Db      27202 AGTGAAAAGAAGATTTTTAAAGTTTATAATTTTTGATTATTTAATTTTATTTATAATTGT 27261

Qy        838 GTGATAGGAGCATTGGTTTATCATTCTATATTTTTTGTTTTTTTTTAGGAGAATATCTTTC 897

Db      27262 GGGAATGAATAAGATATTAATTGTTTTATGTATTTTATTTATATTAATTAATTTGTGTTT 27321

Qy        898 TTATTATA-----TGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTAC 952

Db      27322 TTATTAAAAGTAGTTATATAGAATTTTTTTTAATTTTTGGTAGTAAGTTTAGAAAATGAA 27381

Qy        953 AAGTTGTGTTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGC 1012

Db      27382 GTTTATAGTTATTTTGAATTGGATATATTTTTTGAGTTGATTATTTTTGTAAGTGTAGGA 27441

Qy       1013 TTACAGTTGGAGTTGTTTATGGATTT 1038

Db      27442 ATATAATATTGTTTTTTTATGGTTTT 27467

RESULT 11
```

```
US-10-240-708-38
; Sequence 38, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/240,708
;  CURRENT FILING DATE:  2002-10-03
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 38
;   LENGTH: 5501
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-38

  Query Match             6.8%;  Score 71.8;  DB 10;  Length 5501;
  Best Local Similarity   44.6%;  Pred. No. 0.11;
  Matches  417;  Conservative    0;  Mismatches  507;  Indels   11;  Gaps    3;

Qy         78 TTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTTATGCTTGCTCTT 137
              |||  |  |   |  ||  |||| | ||    | |  |  ||| ||   |  ||  | |
Db       1410 TTTTATTTTATAGTTGAATTATTAATATTTTGGGACGTATATTTTTAGTTTTTTAATTTA 1469

Qy        138 TTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGTCGATTCTTTAAA 197
               ||| |   ||   || |   |  |     | ||| | |   ||||  | ||||  | |
Db       1470 GTTTATTAAAAATTTTTATTATTTTTTAATAAATATTATTAGTTTTTTTGATTAGTAAGG 1529

Qy        198 GTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAAATTATGCTTGTG 257
              |       ||    |  || |  | | ||    |  |  | |    || ||||| || |
Db       1530 TT-----GTGGTGATTGTGTTTTTTTATAATATTAGTGTATATTTTAAGTTATGTTTATT 1584

Qy        258 GGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTAAAGACATT 317
              ||  |||  |   || || |||||  | | |    || || || | ||||
Db       1585 GGAAATTTGTAATTTTTTTTTTTTTGAATTTTTTATTTATATTATTTATTATTTTTTTAA 1644

Qy        318 CGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACTTGTTTAGATT 377
              |       |   | |||   ||      |||||  |  |    | ||| ||||||   |
Db       1645 TGGGGGTAGTTTGTTATTAATTTGGTAGATTTTTTTTTATATATTATTATTTGTTTTTTTA 1704

Qy        378 GGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGATTTTTCTCTT 437
              |||     |   | |    ||| || | ||    | ||||     || ||| ||
Db       1705 TGGATAGTTTTTGTAAAAATTTTGTATATATTTGTATTTTTTTTAATTTATTATTTGAGG 1764

Qy        438 TATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCAG-ATTTTATATT 496
              |||  | |  | | |||    |  |||| | | ||||||  |       |  || || |
Db       1765 AATAAAAGTTAAAAAAATTTTATGTTTGTAATTTAATTATTTGATTAGTTAAGTTGTAGT 1824

Qy        497 CCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGATTCCATACT 556
              |||  |    ||  | ||   | | ||   | || |||||   | |    ||   |
Db       1825 GAGTAGTGATTATTTTTATTTATTTTTTATATTTTTATTTATATATATGATTTGTTTTAG 1884

Qy        557 TTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGGTTTTAATTATT 616
              |||||||  |      |||| | | |||||  ||||     |  ||  | |  ||
Db       1885 TTTTTTTTGTTTATATTTTTTTTTGTTTAATTTTTATTTTAGGTGATATTTAGATTATTT 1944

Qy        617 TCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTACGGGGATGA 676
              |   |  | |  | | |||||||  ||   | |     |    ||     |      |

Db       1945 TGTTTTGAAGAATTATTTTTTTTTAAATAAAAATGTATTTATTTTTGAATTTTAAAAGTT 2004

Qy        677 TATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATGGGGACA-----TT 731
              | | |  ||  |    |  |     ||||| |||   | ||  | | || |     ||
Db       2005 ATAAATGTAGGGAATTGTGGGATTAGTTTTATAAATTTTTTTGTTTTGAGATAGAGTTTT 2064

Qy        732 AAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATTAATTCC 791
              |  ||| ||   ||      |       ||   ||| | |   |     |       ||   |
Db       2065 ATTTTTGTTTCGTAGGTTAGAGTGTAATGGTACGATTTCGGTTTATTGTAATTTTTGTTT 2124

Qy        792 TAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATAGGAGCATT 851
              |   |       |||||||||  |     | |||  |    ||||  | ||||||   |
Db       2125 TTCGGGTTTAAGTGATTTTTTTGTTTTAGTTTTTCGAGTAGTTGGGATTATAGGATTTTG 2184

Qy        852 GGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTATTATATGAGAG 911
              ||||  ||   || |||||||  |||||       ||     ||  |  |     ||
Db       2185 TTATTATGTTTAGTTAATTTTTGTATTTTTAGTAGAGATGGGGTTTTATTATGTTGGTTA 2244

Qy        912 AAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAATTTATAC 971
              || |  |     || |  |     ||   ||| |  |     |||   | ||  |  ||
Db       2245 GGTTGGTTTTAAATTTTTGATTTTAAGTGATTTATTCGTTTTGGTTTATTAAAGTGTTAG 2304

Qy        972 ATTAAACCCTTTTGATGCTTTTAATCGATTGATTT 1006
              |||| |  | |  | |  | |  || |||
Db       2305 ATTATAGGCGTGAGTTATTGTATTTAGTTTATTTT 2339

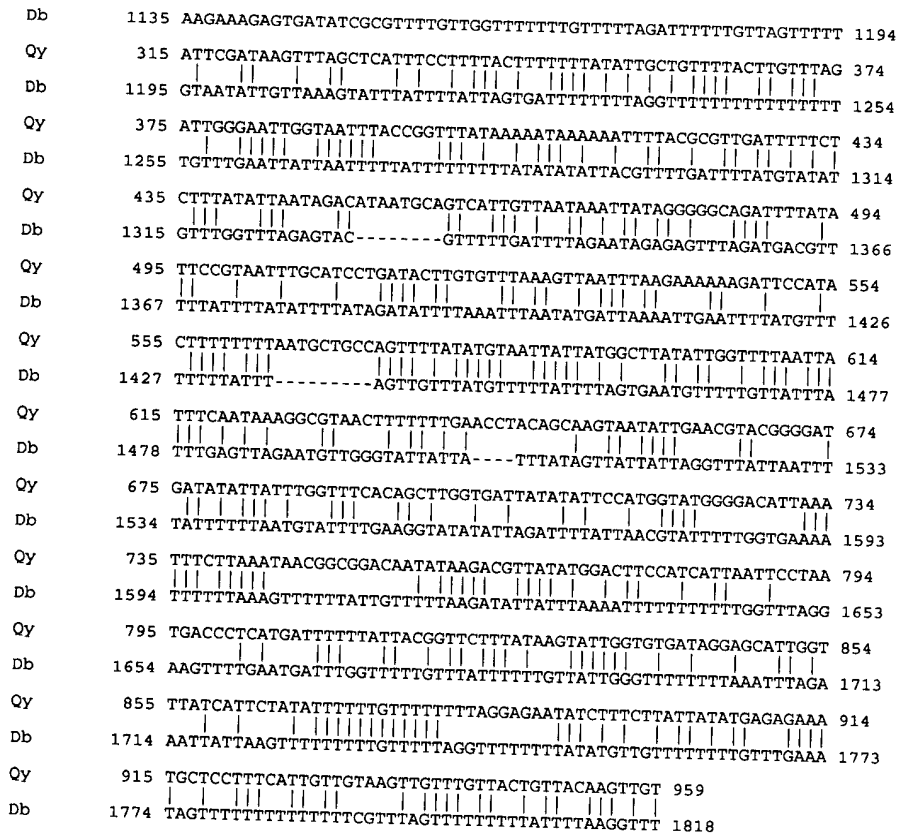

RESULT 12
US-10-240-708-22
; Sequence 22, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/240,708
;  CURRENT FILING DATE:  2002-10-03
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 22
;   LENGTH: 11049
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-22

  Query Match             6.8%;  Score 71.4;  DB 10;  Length 11049;
  Best Local Similarity   46.7%;  Pred. No. 0.13;
  Matches  385;  Conservative    0;  Mismatches  416;  Indels   24;  Gaps    4;

Qy        138 TTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGTCGATTCTTTAAA 197
              | ||  ||||   |||     ||  |||| |||| |  ||     |   |  ||| |  |
Db       1015 TATTAAAGTAGATATTATGTTTGTTGAGATTAAAGATAATTTATATTTTTATTTTGAAGT 1074

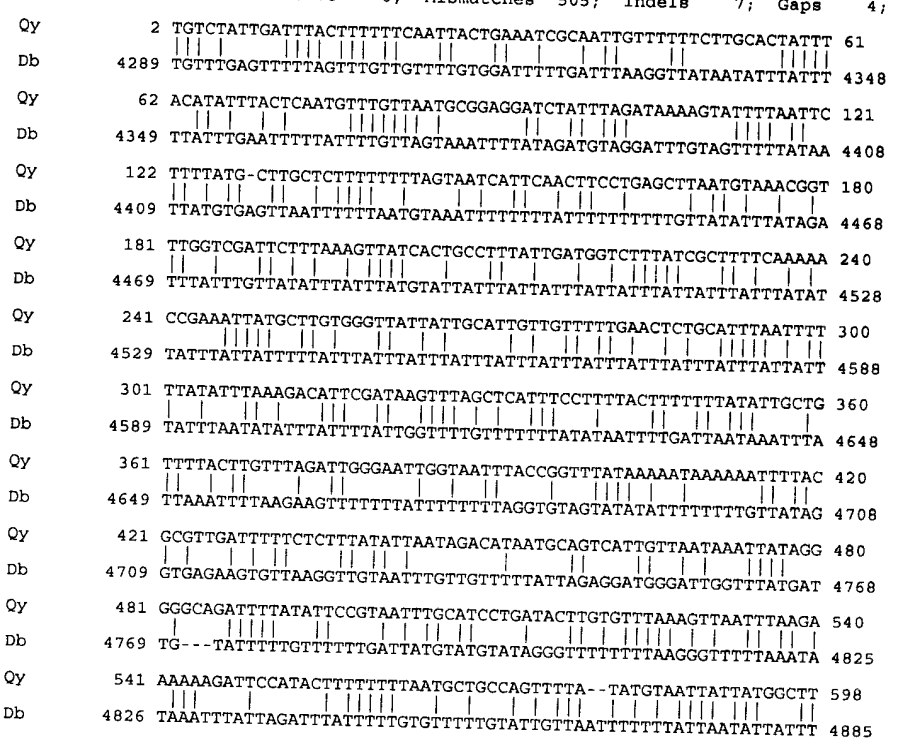
Qy        198 GTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAAATTATGCTTGTG 257
              | || | |    || | |  |  |     | |||| |||||  |  ||  |  ||
Db       1075 TTTTTATATTTTTAAATTTTTTTAAAAAGTGTTTTTTAAAAAATGTTATATTTGTTTAA 1134

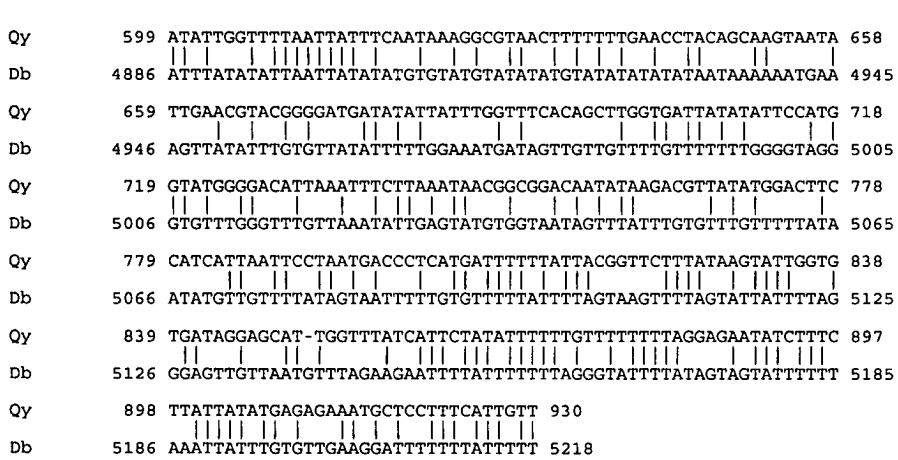
Qy        258 GGTTATTATTGCATT---GTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTAAAGAC 314
              |  | ||   |   ||| | || |   | |   ||| |  |||| | ||
```



RESULT 13
US-11-011-332A-91
; Sequence 91, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
; APPLICANT: Foekens, John
; APPLICANT: Harbeck, Nadia
; APPLICANT: Koenig, Thomas
; APPLICANT: Maier, Sabine
; APPLICANT: Martens, John
; APPLICANT: Model, Fabian
; APPLICANT: Nimmrich, Inko
; APPLICANT: Rujan, Tamas
; APPLICANT: Schmitt, Manfred
; APPLICANT: Lesche, Ralf
; APPLICANT: Dietrich, Dimo
; APPLICANT: Mueller, Volkmar
; APPLICANT: Kluth, Antje
; APPLICANT: Schwope, Ina
; APPLICANT: Hartmann, Oliver
; APPLICANT: Adorjan, Peter
; TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SU
; TITLE OF INVENTION: BREAST CELL PROLIFERATIVE DISORDER PATIENTS
; FILE REFERENCE: 47675-99
; CURRENT APPLICATION NUMBER: US/11/011,332A
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 10/517,741
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: PCT/EP2003/010881
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: DE 10245779.4
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: DE 10300096.8
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: DE 10317955.0
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: PCT/EP2004/014170
; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: EP 03090432.0
; PRIOR FILING DATE: 2003-12-11
; PRIOR APPLICATION NUMBER: EP 04090041.7
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: EP 04090380.9
; PRIOR FILING DATE: 2004-09-30
; PRIOR APPLICATION NUMBER: EP 04090127.4
; PRIOR FILING DATE: 2004-04-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 91
; LENGTH: 6499
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-91

Query Match 6.7%; Score 71; DB 17; Length 6499;
Best Local Similarity 45.1%; Pred. No. 0.14;
Matches 421; Conservative 0; Mismatches 505; Indels 7; Gaps 4;





```
Qy        599 ATATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATA 658
Db       4886 ATTTATATATTAATTATATATGTGTATGTATATATGTATATATATATAATAAAAAATGAA 4945

Qy        659 TTGAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATG 718
Db       4946 AGTTATATTTGTGTTATATTTTTGGAAATGATAGTTGTTGTTTTGTTTTTTTGGGGTAGG 5005

Qy        719 GTATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTC 778
Db       5006 GTGTTTGGGTTTGTTAAATATTGAGTATGTGGTAATAGTTTATTTGTGTTTGTTTTTATA 5065

Qy        779 CATCATTAATTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTG 838
Db       5066 ATATGTTGTTTTATAGTAATTTTTGTGTTTTTATTTTAGTAAGTTTTAGTATTATTTTAG 5125

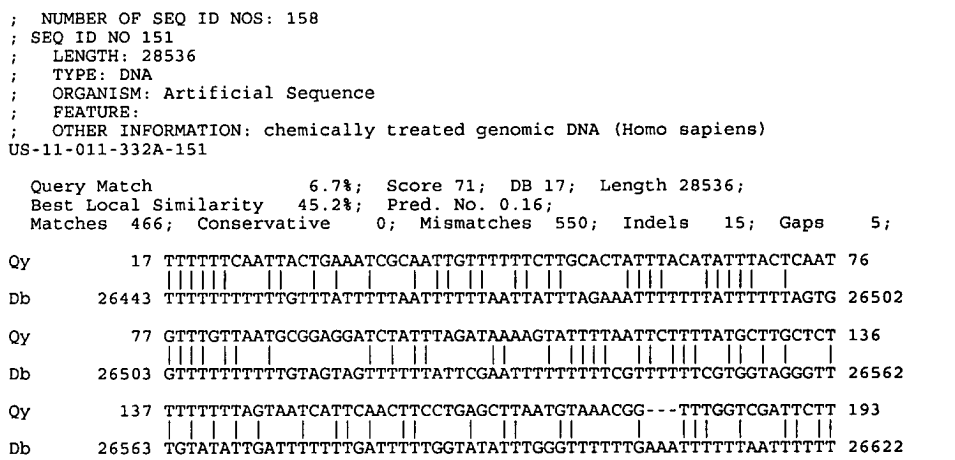
Qy        839 TGATAGGAGCAT-TGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTC 897
Db       5126 GGAGTTGTTAATGTTTAGAAGAATTTTATTTTTTTAGGGTATTTTATAGTAGTATTTTTT 5185

Qy        898 TTATTATATGAGAGAAATGCTCCTTTCATTGTT 930
Db       5186 AAATTATTTGTGTTGAAGGATTTTTTTATTTTT 5218
```

RESULT 14
US-11-011-332A-151
; Sequence 151, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
; APPLICANT: Foekens, John
; APPLICANT: Harbeck, Nadia
; APPLICANT: Koenig, Thomas
; APPLICANT: Maier, Sabine
; APPLICANT: Martens, John
; APPLICANT: Model, Fabian
; APPLICANT: Nimmrich, Inko
; APPLICANT: Rujan, Tamas
; APPLICANT: Schmitt, Manfred
; APPLICANT: Lesche, Ralf
; APPLICANT: Dietrich, Dimo
; APPLICANT: Mueller, Volkmar
; APPLICANT: Kluth, Antje
; APPLICANT: Schwope, Ina
; APPLICANT: Hartmann, Oliver
; APPLICANT: Adorjan, Peter
; TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SU
; TITLE OF INVENTION: BREAST CELL PROLIFERATIVE DISORDER PATIENTS
; FILE REFERENCE: 47675-99
; CURRENT APPLICATION NUMBER: US/11/011,332A
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 10/517,741
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: PCT/EP2003/010881
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: DE 10245779.4
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: DE 10300096.8
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: DE 10317955.0
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: PCT/EP2004/014170
; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: EP 03090432.0
; PRIOR FILING DATE: 2003-12-11
; PRIOR APPLICATION NUMBER: EP 04090041.7
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: EP 04090380.9
; PRIOR FILING DATE: 2004-09-30
; PRIOR APPLICATION NUMBER: EP 04090127.4
; PRIOR FILING DATE: 2004-04-01
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 151
; LENGTH: 28536
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-151

Query Match 6.7%; Score 71; DB 17; Length 28536;
Best Local Similarity 45.2%; Pred. No. 0.16;
Matches 466; Conservative 0; Mismatches 550; Indels 15; Gaps 5;



```
Qy         17 TTTTTTCAATTACTGAAATCGCAATTGTTTTTTCTTGCACTATTTACATATTTACTCAAT 76
Db      26443 TTTTTTTTTTTGTTTATTTTTAATTTTTTAATTATTTAGAAATTTTTTTATTTTTTAGTG 26502

Qy         77 GTTTGTTAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATTCTTTTATGCTTGCTCT 136
Db      26503 GTTTTTTTTTTGTAGTAGTTTTTTATTCGAATTTTTTTTTCGTTTTTTCGTGGTAGGGTT 26562

Qy        137 TTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGG---TTTGGTCGATTCTT 193
Db      26563 TGTATATTGATTTTTTTGATTTTTGGTATATTTGGGTTTTTTGAAATTTTTTAATTTTTT 26622

Qy        194 TAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAAATTATGCT 253
Db      26623 TAGATTTGAGGATGGTAGGTTTTATTTTTTTTATTGTGTGTATATATTTAGAGATATGAA 26682

Qy        254 TGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTAAAGA 313
Db      26683 AATTTATATAGATTGTTTTTAAATTTAGGGTATTTAATAGATGTTTTTTTTTTAGTTCGT 26742

Qy        314 CATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACTTGTTTA 373
Db      26743 TTTTTGATTTGAAATGTTTGTTTGATTTTAATTTGGATATTATTTTTTTTTGTTTTTTTT 26802

Qy        374 GATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGATTTTTC 433
Db      26803 TTTTTAAAGTAGTTTGGATATGTGTGTAAGTGAGTTTAGAATAGTTTTATTTATATTTTT 26862

Qy        434 TCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCAGATTTTAT 493
Db      26863 TATTAAATTGTAAATAAAAGAAGAATTAATGAAGTAGATTGGTATATAGATTGTATTAAG 26922

Qy        494 ATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGATTCCAT 553
Db      26923 AGTTCGAATTTTTAGTTTTTGGATTTTTTATTTAATTTTGGTTGTTATTTATATTGATAG 26982

Qy        554 ACTTTTTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGGTTTTAATT 613
Db      26983 AGTTATTTTAAGTA----GAGGTTTAGAGAAATTTGTATTGTGGGATAATAGGTAAAGTT 27038

Qy        614 ATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTACGGGGA 673
Db      27039 ATAGTAAAAAGTGGAATAATTTTAAAGTTATTTTATTA--GAATGTAAATTGTATTTTTG 27096

Qy        674 TGATATATTATTTGGTTTCACAGCTTGGTGATTATATA-TTCCATGGTATGGGGACATTA 732
Db      27097 GGTTTTGTTCGTAATTATTTAGTTTTAATATATATAGAGTTAGATAGGAAAAAATAGGTT 27156

Qy        733 AATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATTAATTCCT 792
Db      27157 AATATAGTTATTGGTATTAGAGAAGATAAATTTTATGGGTTTTTTAGTGAAAAGAAGATT 27216

Qy        793 AATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATAGGAGCATTG 852
Db      27217 TTTAAAGTTTATAATTTTTGATTATTTAATTTTATTTATAATTGTGGGAATGAATAAGAT 27276

Qy        853 GTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTATTATA-----TG 907
Db      27277 ATTAATTGTTTTATGTATTTTATTTATATTAATTAATTTGTGTTTTTATTAAAAGTAGTT 27336
```

```
Qy        908 AGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAATTT 967
Db      27337 ATATAGAATTTTTTTTAATTTTTGGTAGTAAGTTTAGAAAATGAAGTTTATAGTTATTTT 27396

Qy        968 ATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTTGGAGTTG 1027
Db      27397 GAATTGGATATATTTTTTGAGTTGATTATTTTTGTAAGTGTAGGAATATAATATTGTTTT 27456

Qy       1028 TTTATGGATTT 1038
Db      27457 TTTATGGTTTT 27467
```

RESULT 15
US-10-240-708-17
; Sequence 17, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 17
; LENGTH: 5535
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-17

Query Match 6.6%; Score 69.2; DB 10; Length 5535;
Best Local Similarity 43.7%; Pred. No. 0.25;
Matches 404; Conservative 0; Mismatches 513; Indels 7; Gaps 2;

```
Qy        134 TCTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGTTTGGTCGATTCTT 193
Db       1281 TATTTTTTTTTTTTTTCGTTTTTTGAGACGGAGTTTCGTTTTGTCGTTTAGGTTAGAGTG 1340

Qy        194 TAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAACCGAAATTATGCT 253
Db       1341 TAGTGGTGTAATTTCGGTTTATTGTAATTTTCGTTTTTTGGGTTTAAGTAATTTTTCGTT 1400

Qy        254 TGTGGGTTATTATTGCATTGTTGTTTTTGAACTCTGCATTTAATTTTTTATATTTAAAGA 313
Db       1401 TTAGTTTTTCGAGTAGTTGGGATTATAGGCGTTCGTTATTACGTTCGGTTAATTTTTTTT 1460

Qy        314 CATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTGTTTTACTTGTTTA 373
Db       1461 TTGTATTTTTTAGTAGAGATAAGGTTTTATTATTTTGGTTAGGTTGGTTTTGAATTTTTG 1520

Qy        374 GATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTACGCGTTGATTTTTC 433
Db       1521 ATTTCGTGATTTATTTTTTATCGGTTTTTTAAAGTGTCGAGATTATAGGCGTGAGTATTG 1580

Qy        434 TCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGGGGGCAGATTTTAT 493
Db       1581 TGTTCGGTTTTTTGGGGGAAAATTATATTTAGTGATGAGATTATTAGATTGATTTTTTTT 1640

Qy        494 ATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGAAAAAAGATTCCAT 553
Db       1641 TTTTTTTATTAGGCGTGAGTATTTTTTTTGGTTTTTATATTTTTTAATATTGTAAGAAAT 1700

Qy        554 ACTTT------TTTTTAATGCTGCCAGTTTTATATGTAATTATTATGGCTTATATTGGTT 607
Db       1701 AATATTAAGAATTTAGGATTTTGGATTTTTTTTTTTTTTTTTTTTTTTTTCGTTTTT 1760

Qy        608 TTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATTGAACGTA 667
Db       1761 TTGTTTGTTTGTTTGTTTTGTTTTGTTTGTTTTTGGTTTTATTTTGTTATTTAGGTTGGA 1820

Qy        668 CGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGTATGGGGA 727
Db       1821 GTGTAGTGGTATAATCGTAGTTTATTGTAGTTTTTGTTTTTTGGGTTTAAGTAATTTTTT 1880

Qy        728 CATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCATCATTAA 787
Db       1881 TTTTTTAGTTTTTTGAGTATTTGGGATTATAGGTGTATATTAATATATATAGTTTTTTTT 1940

Qy        788 TTCCTAATGACCCTCATGATTTTTTATTACGGTTCTTTATAAGTATTGGTGTGATAGGAG 847
Db       1941 TTTTTTTTTTTTTTTTAGATGGAGTTTCGTTTTTGTCGTTTAGGTTGGAGTATAATTGCG 2000

Qy        848 CATTGGTTTATCATTCTATATTTTTTGTTTTTTTTAGGAGAATATCTTTCTTATTATATG 907
Db       2001 TGATTTTGGTTTATTATAATTTTTATTTTTCGGGTTTAAGTGATTTTTTTGTTTTAGTTT 2060

Qy        908 AGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTTACTGTTACAAGTTGTGTTAATTT 967
Db       2061 TTTAAAGTGTTGGGATTATAGGTATGAGTTATTTTGTTTAGTTAATTTTTGTGTTTTTTT 2120

Qy        968 ATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTTGGAGTTG 1027
Db       2121 GTAGAG-ATGGGGTTTTGTTATGTTGTTTTGGTTGGTTTTGAATTTTGGAGTTTAAGTAA 2179

Qy       1028 TTTATGGATTTGCAAAAATTAGAT 1051
Db       2180 TTTGTTTATTTTGGTTTTTTAAAT 2203
```

Search completed: May 11, 2006, 08:30:40
Job time : 547.139 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:45:22 ; Search time 4083.52 Seconds
(without alignments)
15660.244 Million cell updates/sec

Title: US-09-423-093B-1_COPY_11821_12945
Perfect score: 1125
Sequence: 1 atgaagattgtttatataat.....gtaataatataattgattga 1125

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1125	100.0	14516	1	AF078736	AF078736 Escherich
	2	1125	100.0	14516	6	BD085554	BD085554 Nucleic a
	3	1125	100.0	14520	6	AX029261	AX029261 Sequence
	4	761.8	67.7	13650	1	AF285969	AF285969 Salmonell
	5	273.6	24.3	13682	1	YETRSATOH	Z47767 Yersinia en
	6	252.8	22.5	12033	1	AY532664	AY532664 Escherich
	7	245.8	21.8	1104	6	AX276712	AX276712 Sequence
	8	245.8	21.8	4604	6	BD248530	BD248530 Virulence
	9	245.8	21.8	4604	6	AR632878	AR632878 Sequence
c	10	245.8	21.8	6927	6	AX276709	AX276709 Sequence
	11	229.8	20.4	12215	1	AY057452	AY057452 Edwardsie
c	12	214.4	19.1	110000	1	CR626927_30	Continuation (31 o
	13	201.8	17.9	1077	6	AR318930	AR318930 Sequence
	14	199	17.7	110000	1	AP006841_17	Continuation (18 o
	15	85.4	7.6	19113	1	CR931708	CR931708 Streptoco
c	16	82	7.3	110000	1	CP000090_07	Continuation (8 of
c	17	82	7.3	110000	1	CP000090_08	Continuation (9 of
	18	81.8	7.3	15155	1	AY208937	AY208937 Escherich
	19	81.8	7.3	17433	1	AY380835	AY380835 Shigella
	20	78.8	7.0	110000	1	AE017340_05	Continuation (6 of
	21	77.6	6.9	370	6	AX100720	AX100720 Sequence
c	22	75.6	6.7	10021	1	AE001736	AE001736 Thermotog
	23	75.4	6.7	110000	1	AE017355_14	Continuation (15 o
	24	75.2	6.7	17618	1	CR931646	CR931646 Streptoco
c	25	74.8	6.6	13528	1	AE007802	AE007802 Clostridi
	26	74.8	6.6	19669	1	AF402095	AF402095 Streptoco
	27	74.8	6.6	20538	1	CR931645	CR931645 Streptoco
	28	74.8	6.6	20856	1	CR931648	CR931648 Streptoco
	29	74.6	6.6	162077	5	BX294656	BX294656 Zebrafish
	30	74.2	6.6	8594	1	AE013034	AE013034 Thermoana
	31	74.2	6.6	14263	1	AF172324	AF172324 Escherich
	32	73.8	6.6	17619	1	CR931647	CR931647 Streptoco
	33	73.8	6.6	110000	1	AE017225_14	Continuation (15 o
	34	73.8	6.6	110000	1	AE017334_14	Continuation (15 o
	35	73.8	6.6	110000	1	CP000001_14	Continuation (15 o
	36	73.8	6.6	110000	1	CP000001_15	Continuation (16 o
	37	73.8	6.6	290525	1	AE017029	AE017029 Bacillus
	38	73.6	6.5	12759	1	AE011518	AE011518 Leptospir
c	39	73	6.5	304923	1	AE017289	AE017289 Leptospir
c	40	72.2	6.4	176147	5	AL929105	AL929105 Zebrafish
	41	72.2	6.4	226847	14	BX664721	BX664721 Danio rer
c	42	72	6.4	349648	1	BX957220	BX957220 Methanoco
c	43	71.2	6.3	110000	1	BA000028_17	Continuation (18 o
c	44	71.2	6.3	110000	1	BA000028_18	Continuation (19 o
c	45	71	6.3	74150	14	BX927093	BX927093 Danio rer

ALIGNMENTS

RESULT 1
AF078736
LOCUS AF078736 14516 bp DNA linear BCT 03-AUG-1999
DEFINITION Escherichia coli O111 O antigen gene cluster, partial sequence.
ACCESSION AF078736 U13629
VERSION AF078736.1 GI:5678713
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 14516)
AUTHORS Bastin,D.A. and Reeves,P.R.
TITLE Sequence and analysis of the O antigen gene (rfb) cluster of Escherichia coli O111
JOURNAL Gene 164 (1), 17-23 (1995)
PUBMED 7590310
REFERENCE 2 (bases 3021 to 9981)
AUTHORS Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C., Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
TITLE Bacterial polysaccharide synthesis and gene nomenclature
JOURNAL Trends Microbiol. 4 (12), 495-503 (1996)
PUBMED 9004408
REFERENCE 3 (bases 1 to 14516)
AUTHORS Wang,L., Curd,H., Qu,W. and Reeves,P.R.
TITLE Sequencing of Escherichia coli O111 O-antigen gene cluster and identification of O111-specific genes
JOURNAL J. Clin. Microbiol. 36 (11), 3182-3187 (1998)
PUBMED 9774562
REFERENCE 4 (bases 3021 to 9981)
AUTHORS Bastin,D.A.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-1997) Microbiology, University of Sydney, Maze Crescent, Sydney, NSW 2006, Australia
REFERENCE 5 (bases 1 to 14516)
AUTHORS Wang,L., Curd,H., Qu,W., Bastin,D.A. and Reeves,P.R.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-1998) Microbiology, University of Sydney, Maze Crescent, Sydney, NSW 2006, Australia
COMMENT On Aug 3, 1999 this sequence version replaced gi:1155016.
FEATURES Location/Qualifiers

```
source          1. .14516
                /organism="Escherichia coli"
                /mol_type="genomic DNA"
                /strain="Stoke W; M92"
                /serotype="O111"
                /db_xref="taxon:562"
gene            <1. .273
                /gene="galF"
CDS             <1. .273
                /gene="galF"
                /codon_start=1
                /transl_table=11
                /product="putative glucose-1-phosphate
                uridylyltransferase"
                /protein_id="AAD46727.1"
                /db_xref="GI:5678714"
                /translation="DLMAVGRYVLSADIWAELEKTAPGAWGRIQLTDAIAELAKKQSV
                DAMLMTGDSYDCGKKMGYMQAFVKYGLRNLKEGAKFRKSIKKLLSE"
gene            739. .1932
                /gene="wbdH"
CDS             739. .1932
                /gene="wbdH"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase"
                /protein_id="AAD46728.1"
                /db_xref="GI:5678715"
                /translation="MLLCCIHINVYYLLLECDMKKIVIIGNVASMMLRFRKELIMNLV
                RQGDNVYCLANDFSTEDLKVLSSWGVKGVKFSLNSKGINPFKDIIAVYELKKILKDIS
                PDIVFSYFVKPVIFGTIASKLSKVPRIVGMIEGLGNAFTYYKGKQTTKTKMIKWIQIL
                LYKLALPMLDDLILLNHDDKKDLIDQYNIKAKVTVLGGIGLDLNEFSYKEPPKEKITF
                IFIARLLREKGIFEFIEAAKFVKTTYPSSEFVILGGFESNNPFSLQKNEIESLRKEHD
                LIYPGHVENVQDWLEKSSVFVLPTSYREGVPRVIQEAMAIGRPVITTNVPGCRDIIND
                GVNGFLIPPFEINLLAEKMKYFIENKDKVLEMGLAGRKFAEKNFDAFEKNNRLASIIK
                SNNDF"
gene            1983. .3101
                /gene="gmd"
CDS             1983. .3101
                /gene="gmd"
                /codon_start=1
                /transl_table=11
                /product="GDP-D-mannose dehydratase"
                /protein_id="AAD46729.1"
                /db_xref="GI:5678716"
                /translation="MNKVALITGITGQDGSYLAELLLEKGYEVHGIKRRASSFNTERV
                DHIYQDSHLANPKLFLHYGDLTDTSNLTRILKEVQPDEVYNLGAMSHVAVSFESPEYT
                ADVDAIGTLRLLEAIRILGLEKKTKFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
                KLYAYWITVNYRESYGMFACNGILFNHESPRRGETFVTRKITRGIANIAQGLDKCLYL
                GNMDSLRDWGHAKDYVKMQWMMLQQETPEDFVIATGIQYSVREFVTMAAEQVGIELAF
                EGEGVNEKGVVVSVNGTDAKAVNPGDVIISVDPRYFRPAEVETLLGDPTNAHKKLGWS
                PEITLREMVKEMVSSDLAIAKKNVLLKANNIATNIPQE"
gene            3174. .3623
                /gene="wbdI"
CDS             3174. .3623
                /gene="wbdI"
                /note="previously known as orf3.4"
                /codon_start=1
                /transl_table=11
                /product="WbdI"
                /protein_id="AAC44881.1"
                /db_xref="GI:1155017"
                /translation="MFITSDKFREIIKLVPLVSIDLLIENENGEYLFGLRNNRPAKNY
                FFVPGGRIRKNESIKNAFKRISSMELGKEYGISGSVFNGVWEHFYDDGFFSEGEATHY
                IVLCYTLKVLKSELNLPDDQHREYLWLTKHQINAKQDVHNYSKNYFL"
gene            3657. .5108
                /gene="manC"
CDS             3657. .5108
                /gene="manC"
                /EC_number="2.7.7.22"
                /note="previously known as RfbM"
                /codon_start=1
                /transl_table=11
                /product="mannose-1-phosphate guanylyltransferase"
                /protein_id="AAC44882.1"
                /db_xref="GI:1155018"
                /translation="MSQCLYPVIIAGGTGSRLWPLSRVLYPKQFLNLVGDSTMLQTTI
                TRLDGIECENPIVICNEDHRFIVAEQLRQIGKLTKNIILEPKGRNTAPAIALAAFIAQ
                KNNPNDDPLLLVLAADHSINNEKAFRESIIKAMPYATSGKLVTFGIIPDTANTGYGYI
                KRSSSADPNKEFPAYNVAEFVEKPDVKTAQEYISSGNYYWNSGMFLFRASKYLDELRK
                FRPDIYHSCECATATANIDMDFVRINEAEFINCPEESIDYAVMEKTKDAVVLPIDIGW
                NDVGSWSSLWDISQKDCHGNVCHGDVLNHDGENSFIYSESSSLVATVGVSNLVIVQTKD
                AVLVADRDKVQNVKNIVDDLKKRKRAEYYMHRAVFRPWGKFDAIDQGDRYRVKKIIVK
                PGEGLDLRMHHHRAEHWIVVSGTAKVSLGSEVKLLVSNESIYIPQGAKYSLENPGVIP
                LHLIEVSSGDYLESDDIVRFTDRYNSKQFLKRD"
gene            5114. .6484
                /gene="manB"
CDS             5114. .6484
                /gene="manB"
                /EC_number="5.4.2.8"
                /note="previously known as RfbK"
                /codon_start=1
                /transl_table=11
                /product="phosphomannomutase"
                /protein_id="AAC44883.1"
                /db_xref="GI:1155019"
                /translation="MNKITCFKAYDIRGRLGAELNDEIAYRIGRAYGEFFKPQTVVVG
                GDARLTSESLKKSLSNGLCDAGVNVLDLGMCGTEEIYFSTWYLGIDGGIEVTASHNPI
                DYNGMKLVTKGARPISSDTGLKDIQQLVESNNFEELNLEKKGNITKYSTRDAYINHLM
                GYANLQKIKKIKIVVNSGNGAAGPVIDAIEECFLRNNIPIQFVKINNTPDGNFPHGIP
                NPLLPECREDTSSAVIRHSADFGIAFDGDFDRCFFFDENGQFIEGYYIVGLLAEVFLG
                KYPNAKIIHDPRLIWNTIDIVESHGGIPIMTKTGHAYIKQRMREEDAVYGGEMSAHHY
                FKDFAYCDSGMIPWILICELLSLTNKKLGELVCGCINDWPASGEINCTLDNPQNEIDK
                LFNRYKDSALAVDYTDGLTMEFSDWRFNVRCSNTEPVVRLNVESRNNAILMQEKTEEI
                LNFISK"
gene            6521. .7444
                /gene="wbdJ"
CDS             6521. .7444
                /gene="wbdJ"
                /note="previously known as orf6.7"
                /codon_start=1
                /transl_table=11
                /product="WbdJ"
                /protein_id="AAC44884.1"
                /db_xref="GI:1155020"
                /translation="MKVLLTGSTGMVGKNILEHDSASKYNILTPTSSDLNLLDKNEIE
                KFMLINMPDCIIHAAGLVGGIHANISRPFDFLEKNLQMGLNLVSVAKKLGIKKVLNLG
                SSCMYPKNFEEAIPEKALLTGELEETNEGYAIAKIAVAKACEYISRENSNYFYKTIIP
                CNLYGKYDKFDDNSSHMIPAVIKKIHHAKINNVPEIEIWGDGNSRREFMYAEDLADLI
                FYVIPKIEFMPNMVNAGLGYDYSINDYYKIIABEIGYTGSFSHDLTKPTGMKRKLVDI
                SLLNKIGWSSHFELRDGIRKTYNYYLENQNK"
gene            7441. .8607
                /gene="wbdK"
CDS             7441. .8607
                /gene="wbdK"
                /note="previously known as orf7.7"
                /codon_start=1
                /transl_table=11
                /product="WbdK"
                /protein_id="AAC44885.1"
                /db_xref="GI:1155021"
                /translation="MITYPLASNTWDEYEYAAIQSVIDSKMFTMGKKVELYEKNFADL
                FGSKYAVMVSSGSTANLLMIAALFFTNKPKLKRGDEIIVPAVSWSTTYYPLQQYGLKV
                KFVDINKETLNIDIDSLKNAISDKTKAILTVNLLGNPNDFAKINEIINNRDIILLEDN
                CESMGAVFQNKQAGTFGVMGTFSSFYSHHIATMEGGCVVTDDEELYHVLLCLRAHGWT
                RNLPKENMVTGTKSDDIFEESFKFVLPGYNVRPLEMSGAIGIEQLKKLPGFISTRRSN
                AQYFVDKFKDHPFLDIQKEVGESSWFGFSFVIKEGAAIERKSLVNNLISAGIECRPIV
                TGNFLKNERVLSYFDYSVHDTVANAEYIDKNGFFVGNHQIPLFNEIDYLRKVLK"
gene            8646. .9911
                /gene="wzx"
CDS             8646. .9911
                /gene="wzx"
                /note="integral membrane protein; previously known as
                orf8.9"
                /codon_start=1
                /transl_table=11
                /product="putative O unit flippase"
```

```
                     /protein_id="AAC44886.1"
                     /db_xref="GI:1155022"
                     /translation="MVLTVKKILAFGYSKVLPPVIEQFVNPICIFIITPLILNHLGKQ
                     SYGNWILLITIVSFSQLICGGCSAWIAKIIAEQRILSDLSKKNALRQISYNFSIVIIA
                     FAVLISFLILSICFFDVARNNSSFLFAIIICGFFQEVDNLFSGALKGFEKFNVSCFFE
                     VITRVLWASIVIYGIYGNALLYFTCLAFTIKGMLKYILVCLNITGCFINPNFNRVGIV
                     NLLNESKWMFLQLTGGVSLSLFDRLVIPLILSVSKLASYVPCLQLAQLMFTLSASANQ
                     ILLPMFARMKASNTFPSNCFFKILLVSLISVLPCLALFFFGRDILSIWINPTFATENY
                     KLMQILAISYILLSMMTSFHFLLLGIGKSKLVANLNLVAGLALAASTLIAAHYGLYAI
                     SMVKIIYPAFQFYYLYVAFVYFNRAKNVY"
     gene            9901. .10953
                     /gene="wzy"
     CDS             9901. .10953
                     /gene="wzy"
                     /codon_start=1

  Query Match             100.0%;  Score 1125;  DB 1;  Length 14516;
  Best Local Similarity   100.0%;  Pred. No. 9.8e-173;
  Matches 1125;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11821 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 11880

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11881 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 11940

Qy        121 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11941 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 12000

Qy        181 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12001 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 12060

Qy        241 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12061 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 12120

Qy        301 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12121 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 12180

Qy        361 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12181 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 12240

Qy        421 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12241 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 12300

Qy        481 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12301 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 12360

Qy        541 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12361 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 12420

Qy        601 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12421 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 12480

Qy        661 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12481 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 12540

Qy        721 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12541 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 12600

Qy        781 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12601 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 12660

Qy        841 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12661 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 12720

Qy        901 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12721 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 12780

Qy        961 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12781 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 12840

Qy       1021 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12841 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 12900

Qy       1081 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 1125
              |||||||||||||||||||||||||||||||||||||||||||||
Db      12901 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 12945


RESULT 2
BD085554
LOCUS       BD085554               14516 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085554
VERSION     BD085554.1  GI:22631164
KEYWORDS    JP 2001522248-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 14516)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 1 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Escherichia coli
            PN   JP 2001522248-A/1
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14516
            FT                   /organism='Escherichia coli'.
FEATURES             Location/Qualifiers
     source          1. .14516
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 1125;  DB 6;  Length 14516;
  Best Local Similarity   100.0%;  Pred. No. 9.8e-173;
  Matches 1125;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11821 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 11880

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11881 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 11940

Qy        121 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      11941 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 12000

Qy        181 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 240

Db      12001 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 12060

Qy        241 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 300

Db      12061 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 12120

Qy        301 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 360

Db      12121 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 12180

Qy        361 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 420

Db      12181 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 12240

Qy        421 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 480

Db      12241 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 12300

Qy        481 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540

Db      12301 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 12360

Qy        541 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 600

Db      12361 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 12420

Qy        601 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 660

Db      12421 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 12480

Qy        661 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 720

Db      12481 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 12540

Qy        721 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 780

Db      12541 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 12600

Qy        781 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 840

Db      12601 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 12660

Qy        841 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 900

Db      12661 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 12720

Qy        901 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 960

Db      12721 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 12780

Qy        961 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 1020

Db      12781 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 12840

Qy       1021 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 1080

Db      12841 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 12900

Qy       1081 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 1125

Db      12901 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 12945

RESULT 3
AX029261
LOCUS       AX029261               14520 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 1 from Patent WO9850531.
ACCESSION   AX029261
VERSION     AX029261.1  GI:10190112
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 1 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .14520
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
ORIGIN

  Query Match           100.0%;  Score 1125;  DB 6;  Length 14520;
  Best Local Similarity  100.0%;  Pred. No. 9.8e-173;
  Matches 1125;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60

Db      11821 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 11880

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120

Db      11881 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 11940

Qy        121 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 180

Db      11941 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 12000

Qy        181 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 240

Db      12001 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 12060

Qy        241 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 300

Db      12061 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 12120

Qy        301 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 360

Db      12121 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 12180

Qy        361 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 420

Db      12181 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 12240

Qy        421 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 480

Db      12241 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 12300

Qy        481 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540

Db      12301 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 12360

Qy        541 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 600

Db      12361 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 12420

Qy        601 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 660

Db      12421 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 12480

Qy        661 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 720

Db      12481 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 12540

Qy        721 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 780

Db      12541 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 12600
```

```
Qy        781 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12601 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 12660

Qy        841 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12661 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 12720

Qy        901 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12721 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 12780

Qy        961 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12781 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 12840

Qy       1021 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12841 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 12900

Qy       1081 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 1125
              |||||||||||||||||||||||||||||||||||||||||||||
Db      12901 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 12945
```

```
RESULT 4
AF285969
LOCUS       AF285969               13650 bp    DNA     linear   BCT 02-AUG-2001
DEFINITION  Salmonella enterica O antigen gene cluster, complete sequence.
ACCESSION   AF285969
VERSION     AF285969.1  GI:15076983
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 13650)
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     The E. coli O111 and S. enterica O35 gene clusters: Gene clusters
            encoding the same colitose containing O antigen are highly
            conserved
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 13650)
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2000) Department of Microbiology, The University
            of Sydney, Sydney, NSW 2006, Australia
FEATURES             Location/Qualifiers
     source          1. .13650
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /serovar="Adelaide"
                     /db_xref="taxon:28901"
                     /note="serogroup: O35"
     gene            1321. .2460
                     /gene="wbdH"
     CDS             1321. .2460
                     /gene="wbdH"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glycosyl transferase WbdH"
                     /protein_id="AAK83009.1"
                     /db_xref="GI:15076984"
                     /translation="MKKIVIIGNIASMMLRFRKELIINLVSQGDDVYCLANDFSPEDL
                     KLLSSWGGKGIKYTLNSKGINPFKDILSVIELKNILDKISPDIVFSYFVKPVIFGTIA
                     SKLSRVPRIVGMIEGLGNAFTYYKGKQSIKTRIIKWMQIFLYKLALPMLNDLVLLNKD
                     DKKDLIERYNIKVNVTILGGIGLNLNDFSYKEPLTDKVTFIFIARLLREKGVFEYIEA
                     AKRVKNIYPNVEFVILGGFENNNPFSLQQHEIELLRKEHDLIYPGHVENVQDWLEKSS
                     VFVLPTSYREGVPRVIQEAMAIGRPVITTNVPGCKDIINDGINGFLVPPFESELLVKK
                     MIYFIDNRSKILEMGLSGRLFAEKNFDAMEKNKTLASIIKANHEF"
     gene            2511. .3629
                     /gene="gmd"
     CDS             2511. .3629
                     /gene="gmd"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-D-mannose dehydratase Gmd"
                     /protein_id="AAK83010.1"
                     /db_xref="GI:15076985"
                     /translation="MNKVALITGITGQDGSYLAELLLEKGYEVHGIKRRASSFNTERV
                     DHIYQDIHLNNPKLYLHYGDLTDTSNLTRILQEVQPDEVYNLGAMSHVAVSFESPEYT
                     ADVDALGTLRLLEAIRILGLEKKTKFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
                     KLYAYWITVNYRESYGMFACNGILFNHESPRRGETFVTRKITRGIANIAQGLDKCLYL
                     GNMDSLRDWGHAKDYVKMQWMMLQQEAPEDFVIATGVQYSVRDFVTMAAKQVGIELVF
                     EGEGINEKGIVASVNGNDAKAVKAGDVIISVDPRYFRPAEVETLLGDPTNAHKKLGWS
                     PEITLCEMVKEMVSSDLAIAKKNVLLKTNNIPTNIPQE"
     gene            3698. .4147
                     /gene="gmm"
     CDS             3698. .4147
                     /gene="gmm"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-mannosyl hydrolase Gmm"
                     /protein_id="AAK83011.1"
                     /db_xref="GI:15076986"
                     /translation="MFIPSDKFREVIKLTPLVSIDILIENENGEYLFGLRNNRPAKNQ
                     FFVPGGRIRKSEFIENAFKRISFAELGKEYDISESLFNGIWEHFYEESFFSDADATHY
                     IVLCYKLKVLKNELNLPADQHCEYIWISEDKISNLNNIHKYSKDYFL"
     gene            4181. .5626
                     /gene="manC"
     CDS             4181. .5626
                     /gene="manC"
                     /codon_start=1
                     /transl_table=11
                     /product="mannose-1-phosphate guanylyltransferase ManC"
                     /protein_id="AAK83012.1"
                     /db_xref="GI:15076987"
                     /translation="MSQCLYPVVIAGGTGSRLWPLSRVLYPKQFLNLVGNSTMLQTTI
                     ARLNGIKCENPIVICNEDHRFIVAEQLRQIDKLTKNIILEPKGRNTAPAIALAAFIAQ
                     KKKPKDDPLLLVLAADHSINNEQAFREAIIKAIPYADAGKLVMFGIIPSVAHTGYGYI
                     KRSVPVDADKDSTAYYVANFVEKPNIQKAQEYIMSEGYYWNSGMFLFRASKYLGELKK
                     FRPDIYNKCESATATANIDMDFVRINETEFINCPEESIDYAVMEKTKDAVVLPIDVGW
                     SDVGSWSSLWDISQKDVHGNVCQGDVINHDSEDSFIYSESSLVATVGVSNLVIVQTKD
                     AVLVADRDKVQNVKNIVDSLKQGGRAEYYMHREVFRPWGKYDAIDQGDRYKVKKIIVK
                     PGEGLDLRMHHHRAEHWIVVSGTAKVSLGNKIKLLVPNESIYIPQGAEYSLENPGVIP
                     LHLIEVSSGDYLESDDIVRFTDRYNSKKFLK"
     gene            5639. .7015
                     /gene="manB"
     CDS             5639. .7015
                     /gene="manB"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphomannomutase ManB"
                     /protein_id="AAK83013.1"
                     /db_xref="GI:15076988"
                     /translation="MNKLTCFKAYDIRGRLGVELDDDIAYRIGRAYGELFKPQTVVVG
                     GDTRLTSESLKKSLSNGLCDAGVNVLDLGMCGTEEIYFSTWHLGVDGGIEVTASHNPI
                     DYNGMKLVTKNARPISSDTGLKDIQTLVELNDFEKINLEKTGNITKYSTRDAYIKHLM
                     GYVNLQKIKKIKIVVNSGNGAAGPVIDDIEKYFLQNNIPVTFVKINNTPDGHFPNGIP
                     NPLLPECREDTSKAVIAHSADFGIAFDGDFDRCFFFDEKGRFIEGYYIVGLLAEVFLE
                     KYPNAKIIHDPRLIWNTIDIVQNYGGKPIMTKTGHAYIKQRMREEDAVYGGEMSAHHY
                     FKDFEYCDSGMIPWILICELLSIRNKNLGELVYNCINAWPASGEINCTLSNPQSEIDK
                     LFFYYKNSAISVDYTDGLSMEFSDWRFNVRCSNTEPVVRLNVESKSNVLLMQEKTREI
                     LSFIGKCV"
     gene            7040. .7963
                     /gene="wbdJ"
     CDS             7040. .7963
                     /gene="wbdJ"
                     /codon_start=1
                     /transl_table=11
                     /product="WbdJ"
                     /protein_id="AAK83014.1"
                     /db_xref="GI:15076989"
                     /translation="MKVLLTGSSGMVGKNILEHASAKDYEILTPTSSDLNLLVIDEIE
                     KFMLANKPECIIHAAGLVGGIHANINRPFDFLANNLLMGLNLVTIAKKLGIKKFLNLG
```

```
                SSCMYPRNFEEAIPEKALLTGELEETNEGYAIAKVAVAKACEYISRENSNYYYKTIIP
                CNLYGKYDKFGDDSSHMIPAVIKKIHHAKANNIPEIEIWGDGNSRREFMYAEDLADLV
                FYIIPKIESIPNMVNAGLGYDYTVNDYYKIIAEEIGYTGTFFHDLTKPTGMKRKLVDI
                SLLNKIGWSSHFNLRDGIRKTYNYYLENQYK"
gene            7960. .9126
                /gene="wbdK"
CDS             7960. .9126
                /gene="wbdK"
                /codon_start=1
                /transl_table=11
                /product="WbdK"
                /protein_id="AAK83015.1"
                /db_xref="GI:15076990"
                /translation="MITYPLASNTWDENEYAAIQSVIDSKMFTMGKNVELYEKNFADL
                FGSKYAVMVSSGSTANLLMIAALFFTHKPKLKRGDEIIVPAVSWSTTYYPLQQYGLKV
                KFVDINKETLNIDVDSLKNAISDKTRAILAVNLLGNPNDFSTIHEIISNKDIILLEDN
                CESMGATFQNKQAGTFGIMGSFSSFYSHHIATMEGGCVVTDDEELYHILLCLRAHGWT
                RNLPKENMVTGTKSDDIFEESFKFVLPGYNVRPLEMSGAIGIEQLKKLPGFISTRRSN
                AQYFVDKFKDHPFLDIQKEIGESSWFGFSLIIKQDTGIERKSLINKLITSGIECRPIV
                TGNFLKNKRVLSYFDYSVHDTVVNAEYIDKNGFFVGNHQIPLLNEIDFLRKVLK"
gene            9178. .10431
                /gene="wzx"
CDS             9178. .10431
                /gene="wzx"
                /codon_start=1
                /transl_table=11
                /product="O antigen flippase Wzx"
                /protein_id="AAK83016.1"
                /db_xref="GI:15076991"
                /translation="MKKFLAFGYSKVLPPIIEQFINPICIFVITPLILNHLGKQSYGN
                WILLITIVSFSQLICGGCSAWIAKVIAEQRATSDYAGKGALRKIAYNLSIIMLSFAIL
                ILLILLSLCVFDVARDNTEFLYAIAICGFFQEVDNLFNGALKGFERFNVSCLFEVITR
                VLWAFIVIYGIQENSVLYYTCLAFILKGTLKYILVCLNISGKFIYPNFDKAGIVNLLN
                ESKWMFLQLTGGISLGLFDRLVIPLILSVNKLSSYVPCLQLVQLMFTISASANQILLP
                MFARMKASNTFPSNCFFKIQLVSLSSALPCLVLLFFGYDILSIWISPAFASENFKLMQ
                ILSISYILLSMMTSFHFLLLGLGKSKLVANLNLIAGLTLAASTLIAAHYGLYAISVVK
                IIYPLFQFYYLYVAFFYFKREKNVY"
gene            10421. .11485
                /gene="wzy"
CDS             10421. .11485
                /gene="wzy"
                /codon_start=1
                /transl_table=11
                /product="O antigen polymerase Wzy"
                /protein_id="AAK83017.1"
                /db_xref="GI:15076992"
                /translation="MSIDFLFSITEIAVVFSCAIYIFVQCLLMHKIYLDKVVLILLCL
                LLCLVIIQLPELNEQGLIDAVKLLLPLLIVFIAFQKPKLCLWIIIALLFFNSTLNFWH
                FKVFDKFSSFPFTFFILLYFLYKLVNDQPPRQRNKKFYVVIFIFILIDVTQSLLISYR
                GQILYSLICVIMLVLRINLRKTIPYIFLTLPFIYVFIMTFIGFNYFNKDVVFFEPTAS
                NIERTGMIYYLVIHANDYIFHGMGTLHFLNEGGQYKMLYGFPSLIPNDPHDFLLRFFI
                SIGMMGALIYHCIFFVFFKNISSLLRQKNSYFVIVSCLLLLQVVLFYTLNPFNAFNRL
                ICGLTIGVIYGFSKSRQIYM"
gene            11451. .12344
                /gene="wbdL"
CDS             11451. .12344
                /gene="wbdL"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase WbdL"
                /protein_id="AAK83018.1"
                /db_xref="GI:15076993"
                /translation="MDFQRVDRYICSESFNTPLISIIIATFNSELVIANCLQSVISQS
                YKNIEVIIMDGGSYDKTLDIARSFKDERIKIVSEKDRGIYDAWNKAVDLSSGDWIAFI
                GSDDVYYHTDAITSLVKGAAISNGAPVVYGRTAHEGPNKEISGFSGSEWYNLKGFKFN
                YYKCNLPLPIMSAIYSRDFFKDQRFDIKLKIVADADWFLRCFIKWRKENSPYFIDDIT
                PVVRMGYGGVSTDISSQIKTTRESFIVRKKNNISCFNMQLILRYIKIMVMVSIKNIFG
                NNIYNSMHNGYHFLKKIKNKI"
gene            12341. .13465
                /gene="wbdM"
CDS             12341. .13465
                /gene="wbdM"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase WbdM"
                /protein_id="AAK83019.1"
                /db_xref="GI:15076994"
                /translation="MKILYVITGLTCGGAEHLMMQLADQMLLRGHDVNIICLSGISEI
                QPTQKHNIHYIRMEKNVRGFVKALFQVRKIITVIKPDIIHSHMFHANIFSRIIRILTP
                SVPLICTAHNKNEGGYARMLCYRLSDFLASITTNVSQEAVQEFIAKKAAPKNKIIEVP
                NFVNTTNFSFDLKARKERRNFFGIKDNTLLLLAVGRLVDAKDYPNLLNAINYLTLSES
                TGFDNFILFIAGDGVLKSKLVDLVNKLNLADKVFFLGRRSDIKELMCAADLFVLSSQW

  Query Match           67.7%;  Score 761.8;  DB 1;  Length 13650;
  Best Local Similarity  79.8%;  Pred. No. 5.7e-114;
  Matches  898;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||||||| ||||| |||||||||||  | |||||  ||  ||  || |||||  |||||||  |
Db      12341 ATGAAGATTCTTTATGTAATAACCGGTTTAACTTGCGGCGGGGCTGAACATCTTATGATG 12400

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
              ||| |||| ||||||||||||  |||| ||  ||||| ||||||||||||||||||||  ||||
Db      12401 CAGCTAGCTGACCAAATGTTGCTACGTGGCCATGACGTTAATATTATTTGTCTAAGTGGT 12460

Qy        121 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 180
              |||||||||  ||||  ||||||||     ||||||||||||  ||  ||||||  ||||||
Db      12461 ATATCTGAGATTCAGCCGACACAAAAACACAATATTCATTATATTCGTATGGAAAAAAAT 12520

Qy        181 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 240
              | ||| | ||| | |  |||||  |||||||  | |||||||||    |  ||||||||
Db      12521 GTCAGAGGATTTGTAAAGGCTTTGTTTCAAGTCAGAAAAATAATAACTGTTATAAAGCCA 12580

Qy        241 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 300
              |||||||||||||| ||||||||||||||||||||||||||||||| |||||||||  ||
Db      12581 GATATAATACATAGCCATATGTTTCATGCTAATATTTTTAGTCGTATTATTAGGATACTA 12640

Qy        301 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 360
              | |||| |||| ||  | |||||||| || |||||  ||||||||||||||  |||| |||
Db      12641 ACTCCATCGGTACCATTAATATGTACAGCTCACAATAAAAATGAAGGTGGATATGCTAGG 12700

Qy        361 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 420
              ||| |||||||||||||  || |||||||| || || ||  ||  ||||  |||||  ||||
Db      12701 ATGCTTTGTTATCGACTAAGCGATTTTTTGGCATCGATCACAACAAACGTAAGTCAAGAA 12760

Qy        421 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 480
              |||||  |||||  ||||||||||  |||||  ||||||||||||||||  | ||  | ||
Db      12761 GCTGTGCAAGAATTTATAGCAAAGAAGGCAGCACCTAAAAATAAAATAATCGAAGTCCCC 12820

Qy        481 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540
              ||||||  |||||||  | ||| ||  |||  ||| |||| |  ||||||  |||  || ||
Db      12821 AATTTTGTTAATACGACTAATTTCAGTTTCGATCTTAAAGCTAGAAAGGAAAGAAGAAAT 12880

Qy        541 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 600
              ||||  | | |||||  | |||   | | |||||  ||  ||  |||  ||||||||||  |||
Db      12881 TTTTTTGGTATAAAAGATAATACTTTACTCCTGCTTGCTGTTGGACGACTTGTTGACGCA 12940

Qy        601 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 660
              |||||  |||||||| ||||||||||||||||||  |||| | ||||||  |||  |
Db      12941 AAAGATTATCCGAATTTATTAAATGCAATAAATTATTTAACTCTTTCGGAAAGCACGGGA 13000

Qy        661 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 720
              |   ||  ||||||||||  |||||||  ||  |||||||  |||||  ||  |||||||  |||
Db      13001 TTCGATAATTTTATTTTATTTATTGCCGGTGATGGCGTATTAAAAAGTAAATTAGTAGAT 13060

Qy        721 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 780
              ||  |||  | ||||  |||||||  ||||||||||||  ||  ||||||  |||||||||  |||
Db      13061 TTAGTTAATAAATTAAATCTTGCGGATAAAGTTTTTTTTCTGGGGCGAAGAAGTGACATT 13120

Qy        781 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 840
              ||||||||||||||||||||  |||||  ||||||||  | ||||||||||||||||||
Db      13121 AAAGAATTAATGTGTGCTGCCGATCTCTTTGTTCTGAGTTCTCAATGGGAAGGTTTTGGT 13180

Qy        841 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 900
              |||||||||||||| ||||||||||  |||||  |  ||||||||  || ||||||||||
Db      13181 CTCGTTGTTGCAGAGGCTATGGCGTGCGAACGGCTTGTTGTTGCAACAGATTCTGGTGGT 13240
```

```
Qy        901 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 960
              |||| |||||| ||||| ||||||||||||||| |||||||| |||||||||||||| ||
Db      13241 GTTAGAGAAGTTGTTGGGCCTCATAATGATGTCATCCCTGTTAGTAATCATATTCTGCTG 13300

Qy        961 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 1020
              || |||||||| || |    | | || || ||||| ||  | | ||| || || | | ||
Db      13301 GCGGAGAAAATTGCCGGTGTATTGAAGATGGATGAAAAAACGAAAAATATGATTGCTTTG 13360

Qy       1021 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 1080
              ||||||||| || | ||||||||  ||||||| ||| |||  || || ||| | ||||||
Db      13361 AAAAATAGACAACACATTGTTTCTCATTTTTCGATTCAAAATATCGTTAGTAATTGGGAG 13420

Qy       1081 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 1125
              | ||||||| ||||||||||   ||| | ||| |  |  ||||
Db      13421 TGTTTATATTATAAATATTCCAGATGTACTGATAAAGCTAGTTGA 13465
```

```
RESULT 5
YETRSATOH
LOCUS       YETRSATOH              13682 bp    DNA     linear   BCT 18-APR-2005
DEFINITION  Yersinia enterocolitica adk gene, hemH gene, wzx gene, wbcK gene,
            wbcL gene, wbcM gene, wbcN gene, wbcO gene, wbcP gene, wbcQ gene
            gne gene and gsk gene (partial).
ACCESSION   Z47767 X63827
VERSION     Z47767.1  GI:633689
KEYWORDS    adenylate kinase; adk gene; ferrochelatase; glycosyl transferase;
            gne gene; gsk gene; hemH gene; inosine-guanosine kinase;
            UDP-N-acetylglucosamine-4-epimerase.
SOURCE      Yersinia enterocolitica
  ORGANISM  Yersinia enterocolitica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Yersinia.
REFERENCE   1  (bases 1 to 13682)
  AUTHORS   Skurnik,M., Venho,R., Toivanen,P. and al-Hendy,A.
  TITLE     A novel locus of Yersinia enterocolitica serotype O:3 involved in
            lipopolysaccharide outer core biosynthesis
  JOURNAL   Mol. Microbiol. 17 (3), 575-594 (1995)
   PUBMED   8559076
REFERENCE   2  (bases 1 to 13682)
  AUTHORS   Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C.,
            Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
  TITLE     Bacterial polysaccharide synthesis and gene nomenclature
  JOURNAL   Trends Microbiol. 4 (12), 495-503 (1996)
   PUBMED   9004408
REFERENCE   3  (bases 1 to 13682)
  AUTHORS   Skurnik,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-1992) M. Skurnik, Turku University, Dept. of
            Medical Microbiology, Kiinamyllynkatu 13, SF-20520 Turku, FINLAND
COMMENT     On Oct 18, 2002 this sequence version replaced gi:48595.
FEATURES             Location/Qualifiers
     source          1. .13682
                     /organism="Yersinia enterocolitica"
                     /mol_type="genomic DNA"
                     /strain="6471/76 (serotype O:3)"
                     /db_xref="taxon:630"
     gene            94. .753
                     /gene="adk"
     RBS             94. .97
                     /gene="adk"
                     /citation=[1]
     CDS             109. .753
                     /gene="adk"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="adenylate kinase"
                     /protein_id="CAA87696.1"
                     /db_xref="GI:633690"
                     /db_xref="GOA:P43412"
                     /db_xref="InterPro:IPR000850"
                     /db_xref="InterPro:IPR006259"
                     /db_xref="InterPro:IPR007862"
                     /db_xref="UniProt/Swiss-Prot:P43412"
                     /translation="MRIILLGAPGAGKGTQAQFIMEKYGIPQMCTGDMLRAAVKAGSE
                     LGLKAKEIMDAGKLVTHELVIALVKERITQDDCRDGFLLHGFPRTIPQADAMKEAGIK
                     VDYVLEFDVPDELIVDRIVGRRVHAASGRVYHIKFNPPKVEDKDDVTGEELTIRKDDQ
                     EATVRKRLVEYHQQTAPLVSYYRKEADAGNTQYFKLDGTRKVAEVSAELATILG"
     gene            869. .1807
                     /gene="hemH"
     RBS             869. .873
                     /gene="hemH"
                     /citation=[1]
     CDS             881. .1807
                     /gene="hemH"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="ferrochelatase"
                     /protein_id="CAA87697.1"
                     /db_xref="GI:633691"
                     /db_xref="GOA:P43413"
                     /db_xref="UniProt/Swiss-Prot:P43413"
                     /translation="MKQSKLGVLMVNLGTPDAPTPQAVKRYLAEFLSDRRVVDTSPWL
                     WWPLLRGVILPIRSPRVAKLYQSVWMDEGSPLLVYSRRQQKALAERMPEIPVELGMSY
                     GSPNLPDAIDKLLAQGVTKLVVLPLYPQYSCSTSAAVWDAVARILKGYRRLPSISFIR
                     DYAEHPAYISALKQSVENSFVQHGKPDRLVLSFHGIPKRYAQLGDDYPQRCEDTSRAL
                     RAEIALPAEQIMMTYQSRFGREPWLTPYTDETLKSLPSQGVKHIQLICPGFSADCLET
                     LEEIKEQNREVFIHAGGEKFEYIPALNDDEGP"
     misc_recomb     2035. .2092
                     /standard_name="JUMPstart sequence"
                     /citation=[1]
                     /organism="Yersinia enterocolitica"
     gene            join(2161. .2204,2212. .3468)
                     /gene="wzx"
     -35_signal      2161. .2166
                     /gene="wzx"
                     /citation=[1]
     -10_signal      2182. .2187
                     /gene="wzx"
                     /citation=[1]
     RBS             2202. .2204
                     /gene="wzx"
                     /citation=[1]
     CDS             2212. .3468
                     /gene="wzx"
                     /function="lipopolysaccharide O-unit flippase"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA87698.1"
                     /db_xref="GI:633692"
                     /db_xref="GOA:Q56913"
                     /db_xref="InterPro:IPR002797"
                     /db_xref="UniProt/TrEMBL:Q56913"
                     /translation="MSVLRRNIFSLLLLQGSNYIIPLLTLPYLTRVLGVEGFGVYSLT
                     LSLAQYFVILIDFGFNLSASKKIAEHQDDPEYISKIFFETLISKSILCLVSIVIVVLL
                     LATNSHAVIQRELIYTVLMLIGTTLMPVWFFQGIEKLSIVTNLMIVAKLSMLPLFFIF
                     VDNDADVKYRRNDSIINQPSGWIIAIIYIYRKKLVRVVDFSNLRVSYTLKDSFPIFCA
                     TLSISLYTMSTTIIIGIFSNVYEVSIFTAADRVKGAILGVFLVLGNAFYPRINSLLVN
                     KKEEAYALIKKIFYWQGALCIIITIFVIAFSKLITRIMFGHEYSEVSTLLILFSPVYL
                     LVIQSTVLGNYILLPHGHKTSYTILPMVSAVIHIPLCAYLASKYGAWGGVLSIIVIET
                     LSFICLILILKRKKLLFEVLSRNKNG"
     gene            3450. .3452
                     /gene="wbcK"
     RBS             3450. .3452
                     /gene="wbcK"
                     /citation=[1]
     gene            3461. .4417
                     /gene="wbcK"
     CDS             3461. .4417
                     /gene="wbcK"
                     /function="putative glycosyltransferase"
                     /citation=[1]
                     /codon_start=1
```

```
                     /transl_table=11
                     /protein_id="CAA87699.1"
                     /db_xref="GI:633693"
                     /db_xref="InterPro:IPR001173"
                     /db_xref="UniProt/TrEMBL:Q56914"
                     /translation="MDNKKLLSLIIPVFNAEQYLDACLSSVFVQWDNTLEIIVINDGS
                     TDNSANIIQKYSEKYDFVYISQENTGISVVRNMGLSVSTGEYITFLDADDIWCNEIYS
                     SIKKIVLENSPDGMVFNYSEVLDDKEHVFELIKVNKFTVDRLEHVKLKIAESEMFYVW
                     RCVFKKNIFSGMVFDVGRRFEDQLLLPILIDKCKSIFECKDLIVKYRQISSSITKNLN
                     ISDLDDSEFGLIRFRDKYAEYKNKYWAVVLASVFMSHISKCARIYHFNKIRALDSFNK
                     SYAIVSMKPILHSGKLKPILYYIVKYKLFYRLVSAVEKEVRK"
gene                 4409. .4412
                     /gene="wbcL"
RBS                  4409. .4412
                     /gene="wbcL"
                     /citation=[1]
gene                 4419. .5297
                     /gene="wbcL"
CDS                  4419. .5297
                     /gene="wbcL"
                     /note="putative glycosyltransferase"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA87700.1"
                     /db_xref="GI:633694"
                     /db_xref="InterPro:IPR001173"
                     /db_xref="UniProt/TrEMBL:Q56915"
                     /translation="MISVTVFTPTFNRAHVLKRCYLSILEQDRDDIEWLIVDDGSTDN
                     TAEVVDSFKIENKLNIKYIYQDNSGKQAAWNKAVENASGEYFIGLDSDDAFIAGSINK
                     LLSMNAVFDDKEIIGIRAISVSSETLKPNNYYLSNEDKKSSWFDEFSSGIRGERIDFF
                     KTELLRKYLYPVASGINFIPEIWFYSTVAKEYCFYYSSIPVRIFFDDDKNNRLSKSSI
                     KKNAIGHYISRKALLENMPKTIWLYRPVDFIKSVLRLNQTLTMIDNKYIDHSKSSFLV
                     KIVALPSIIINKIKRR"
RBS                  5295. .5299
                     /citation=[1]
gene                 5306. .6382
                     /gene="wbcM"
CDS                  5306. .6382
                     /gene="wbcM"
                     /note="putative glycosyltransferase"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA87701.1"
                     /db_xref="GI:633695"
                     /db_xref="GOA:Q56916"
                     /db_xref="InterPro:IPR001296"
                     /db_xref="UniProt/TrEMBL:Q56916"
                     /translation="MKKIAFFGGDISHTGGTERVSITLANYLVKNGYQVVIISLSGNK
                     PPKFPVDESIKLVSLFDEKRRFSLAYFSVVVRLRRILVDESIDILIDVDTMLALFSTT
                     ALLGTKIKHISWEHFNYKSNLNIKSRKLARKIAAKYSDAVVTLQRRSWLLARRKQISR
                     KNNIYIQPFAFWNPKIKLTKRNSQIVLALGRLTNIKGFDLLLDIWVKVEEQSSNWKLI
                     IAGDGEDKNLLLDKIKEFNLKNVELLPSTLHVSDLYDQSAIYAMTSRFEGFPMVLLEA
                     KASGLPIIAYDCDTGPSELIINNEDGLLIPFSDSNAFARQLILLMNDDDLRESMSLRS
                     LKNAEKYKIEVAIGDKWKSLIEKI"
gene                 6575. .7609
                     /gene="wbcN"
CDS                  6575. .7609
                     /gene="wbcN"
                     /function="putative glycosyltransferase"
                     /note="alternative start codon 6599. .6601"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA87702.1"
                     /db_xref="GI:633696"
                     /db_xref="GOA:Q56917"
                     /db_xref="InterPro:IPR001296"
                     /db_xref="UniProt/TrEMBL:Q56917"
                     /translation="MYEAGHNVMIISLTGETLVRPNDGIQLNELKLDKAPFSLFKGLF
                     EVKKIIKKFKPDIVHSHMFHANLFARILRVFTKIPALICTAHNTNEGSSLRMLAYKYT
                     DKLASLSTNVSQDAVDSFIHKGASSTGRMIAVSNGIDASQFDFSMDERKVKRSELGIF
```

```
NDTPIILSVGRLTEAKDYPNLLTAFSLLIKDNSLQSFPQLFIVGTGHLDGYLKNMSKE

 Query Match           24.3%;  Score 273.6;  DB 1;  Length 13682;
 Best Local Similarity 54.9%;  Pred. No. 5.5e-35;
 Matches  585;  Conservative    0;  Mismatches  474;  Indels    7;  Gaps    2;

Qy          2 TGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACGC 61
              ||||  |  | || || || || ||||||  |   || ||    |||| || |  | |
Db       6505 TGAATTTATTATACATTATTACAGGGCTTGGTATGGGGGGCTGAGAACAACTGTTTGATC 6564

Qy         62 AGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGTA 121
               ||    | |||||| |  |  || ||| |||| |  |||||||  ||||||||
Db       6565 GCCTA----ATAAAATGTATGAAGCAGGTCATAATGTAATGATTATTTCGCTAACTGGCG 6620

Qy        122 TATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAATT 181
              | |   || |  || |   |   ||| |  |  || |  | || | || ||   |
Db       6621 AAACATTAGTTAGACCCAATGATGGCATTCAATTAAATGAATTAAAGCTAGACAAGGCTC 6680

Qy        182 TTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCAG 241
              |  ||| | |||| ||  |||||| ||||||||||||||||          || || || |
Db       6681 CTTTTAGCCTATTTAAAGGGTTATTTGAAGTAAAAAAAATAATAAAAAAATTCAAACCTG 6740

Qy        242 ATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTGA 301
              | ||  | |||||| |||||||| |||||||  | ||     |  || || || |  |
Db       6741 ACATTGTGCATAGCCATATGTTCCATGCTAACTTATTCGCCAGGATTCTTCGGGTTTTTA 6800

Qy        302 TTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGGA 361
                |        |  | |||||||| |||||  ||||  |||||||  |  |    || |
Db       6801 CGAAAATACCAGCTTTAATATGTACAGCACATAACACCAATGAAGGAAGTTCATTGAGAA 6860

Qy        362 TGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAGG 421
              ||||    ||   |  |  ||   | |||||| |  |  || ||||| ||  |||| |
Db       6861 TGTTAGCCTACAAATATACCGACAAACTTGCTTCTCTGTCGACCAATGTTAGCCAAGATG 6920

Qy        422 CTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCGA 481
              | ||| |    || ||   | |   ||  ||  ||     || |||  | |    || | |
Db       6921 CAGTTGATTCCTTCATTCATAAAGGTGCCTCAAGTACTGGTAGAATGATTGCTGTTTCTA 6980

Qy        482 ATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGATG 541
              ||  |||| || |   | | ||||||||     ||  |||  || |||  ||  ||    |
Db       6981 ATGGTATTGATGCTTCTCAGTTTGATTTCTCAATGGATGAGAGGAAGGTAAAACGGTCAG 7040

Qy        542 CTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCAA 601
                |  | |     |  ||| || || | |||   |||||  | ||     |||||||
Db       7041 AGCTAGGTATATTTAATGATACTCCAATAATTCTCAGCGTAGGGCGGCTAACAGAAGCAA 7100

Qy        602 AAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAATT 661
              |||| ||||| |||||| |||  ||  |   |  |||| || ||  | | |
Db       7101 AAGATTATCCAAACTTACTAACAGCTTT---TTCTTTGCTTATTAAAGATAATAGCTTAC 7157

Qy        662 GTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATT 721
              | | |  |   ||  |||||| |||    ||      |  | |   |  |||
Db       7158 AAAGTTTTCCTCAGTTATTTATTGTTGGAACCGGGCATCTTGACGGTTATCTTAAGAATA 7217

Qy        722 TGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTA 781
              ||       | ||  | |||   | | ||    || || || ||||||  ||||||
Db       7218 TGTCCAAAGAGTTTGGTATTGATAAATATGTAACATTATTTGGTCAAAGAGATGATATAC 7277

Qy        782 AAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTC 841
              | |||||||||||||||| ||| | |||||  |   |||||| ||||||| |||   |
Db       7278 TACAATTAATGTGTGCTGCTGATATCTTTGTGCTCTCTTCTGAATGGGAAGGCTTTCCGC 7337

Qy        842 TCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAG 901
              | ||| || | |||||  |||||  ||  |          | |||||  || ||| | |||||
Db       7338 TTGTTATTACTGAAGCGATGGCATGCAAGAAAATTATCGTTGCAACTGATGCCGGTGGTA 7397

Qy        902 TTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTGG 961
              | |  ||||   | ||   |     |  || | ||  | |  | | || |   ||
Db       7398 TAACCGAAGCATTAGGCGATTGCGGTTCGATTGTACCGATAAAAGACCCTAATAGTTTAT 7457

Qy        962 CAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATGA 1021
              |  |  |||   | | |  | ||| ||  ||| || | || | | ||| ||||||  |
Db       7458 CTCAAGCAATAAATAAAATGATAAAATTATCTGACAATGAAAAAGAGATATTAGGTAATA 7517
```

```
Qy       1022 AAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGT 1067
              ||   ||||||   ||| |    |     || ||| | | |||| |
Db       7518 AAGCGAGAGAACGAATTATACAAACAAACTCCATTGAGAAGATAAT 7563


RESULT 6
AY532664
LOCUS       AY532664               12033 bp    DNA     linear   BCT 30-MAR-2004
DEFINITION  Escherichia coli serotype O103 O-antigen gene cluster, complete
            sequence.
ACCESSION   AY532664
VERSION     AY532664.1  GI:45644916
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 12033)
  AUTHORS   Fratamico,P.M., Briggs,C.E., Debroy,C., Strobaugh,T.P. Jr. and
            Chen,C.-Y.
  TITLE     DNA Sequence of the Escherichia coli O103 O Antigen Gene Cluster
            and Detection of Enterohemorrhagic E. coli O103 by PCR
            Amplification of the wzx and wzy Genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 12033)
  AUTHORS   Fratamico,P.M., Briggs,C.E., Debroy,C., Strobaugh,T.P. Jr. and
            Chen,C.-Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-2004) Eastern Regional Research Center,
            Agriculture Research Service, U. S. Department of Agriculture, 600
            E. Mermaid Ln., Wyndmoor, PA 19038, USA
FEATURES             Location/Qualifiers
     source          1. .12033
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /serotype="O103"
                     /db_xref="taxon:562"
     gene            107. .1183
                     /gene="rmlB"
     CDS             107. .1183
                     /gene="rmlB"
                     /note="RmlB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dTDP-glucose 4,6-dehydratase"
                     /protein_id="AAS73163.1"
                     /db_xref="GI:45644917"
                     /translation="MKILVTGGAGFIGSAVVRHIINNTQDSVVNVDKLTYAGNLESLA
                     DVSDSERYVFEHADICDSVAMARIFAQHQPDAVMHLAAESHVDRSITGPAAFIETNIV
                     GTYVLLEAARNYWSALDGDKKNSFRFHHISTDEVYGDLPHPDEVNNTEELPLFTETTA
                     YAPSSPYSASKASSDHLVRAWKRTYGLPTIVTNCSNNYGPYHFPEKLIPLVILNALEG
                     KGLPIYGKGDQIRDWLYVEDHARALYTVVTEGKAGETYNIGGHNEKKNIDVVLTICDL
                     LDEIVPKEKSYREQITYVADRPGHDRRYAIDAEKIGRELGWKPQETFESGIRKTINWY
                     LNNNDWCRRVQDGSYQRERLGVIR"
     gene            1180. .2052
                     /gene="rmlA"
     CDS             1180. .2052
                     /gene="rmlA"
                     /note="RmlA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative D-glucose-1-phosphate
                     thymidylyltransferase"
                     /protein_id="AAS73164.1"
                     /db_xref="GI:45644918"
                     /translation="MKGIILAGGSGTRLYPITMGVSKQLLPIYDKPMIYYPLSVLMLA
                     GIQDILIITTPEDQQGFMRLLGDGSQFGINLTYAIQPSPDGLAQAFIIGENFIGNDTV
                     CLMLGDNIFFGQGFTPKLRDAAERTSGATVFGYQVMDPERFGVVEFDRNFKALSIEEK
                     PANPKSNWAVTGLYFYDNNVIDIAKCIKPSKRGELEITAVNEVYLRNNNLTVELLGRG
                     FAWLDTGTHDSLIEAGSFVETVQKRQGMMVACPEEIAWRNGWLNHDELQAKGEFLKKN
                     HYGQYLLRLIKSHE"
     gene            2045. .2452
                     /gene="wbtA"
     CDS             2045. .2452
                     /gene="wbtA"
                     /note="WbtA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dTDP-6-deoxy-3,4-keto-hexulose
                     isomerase"
                     /protein_id="AAS73165.1"
                     /db_xref="GI:45644919"
                     /translation="MNNNVKLIHFDKKGDNRGFLVVVEQNKDIPIEIKRVYYMFDTGS
                     DVRRGYHAHKKLTQIAIPVKGSCKFLLDDGKSVEHVTLDSPSVGLIIEPMIWHEMYDY
                     SSDCVLMVLADDFYDESDYIRNYADFIREMTNA"
     gene            2445. .2978
                     /gene="wbtB"
     CDS             2445. .2978
                     /gene="wbtB"
                     /note="WbtB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative butyryltransferase"
                     /protein_id="AAS73166.1"
                     /db_xref="GI:45644920"
                     /translation="MRNLSLFTKNTKLIPVEINDAEFIYSLRVDSRLNKHISTINGTV
                     IEQAEWIEDYKKREENNEEFYFIIKRTDNDEKIGTIRLYNITKDDGFCWGSWILNDDK
                     MKSSAIESAYLIYKFGFLHKNYKYSYFQVDKKNKEVIAFHKKTGATIIKEDDVNFHFE
                     FYVENFIRLREKYSKYF"
     gene            2975. .4096
                     /gene="wbtC"
     CDS             2975. .4096
                     /gene="wbtC"
                     /note="WbtC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative aminotransferase"
                     /protein_id="AAS73167.1"
                     /db_xref="GI:45644921"
                     /translation="MRGKKMIDFLNLKRINAQYEMELKDACSRVIDSGWYIMGTELKT
                     FEEKFANWCGVKHVIGVANGLDALILVLRAWKEMGKLNDGDEIIVPANTYIASVLAIT
                     ENNLVPVLTEPDPLTFNLSAGGIRKVITGKTKAILPVHLYGQISPMAEIMEIAAEFGL
                     LVLEDCAQAHGSSIDGRKAGSWGHAAGFSFYPGKNLGALGDAGAITTSDDQLADILTA
                     IRNYGSHEKYKNKYQGLNSRLDEMQAAFLNVKLKYLDRENKLRRAIAKKYIEEIKNPL
                     ISLPVIAGNEMSHVWHLFVIRTSERDSLQKHLKTEGITTLVHYPIAPHNQEAYKEYSF
                     LSLPLTERIHREVLSLPMDPTMSIDDINMVIEAINRWKS"
     gene            4084. .5349
                     /gene="wzx"
     CDS             4084. .5349
                     /gene="wzx"
                     /note="Wzx"
                     /codon_start=1
                     /transl_table=11
                     /product="putative O-antigen flippase"
                     /protein_id="AAS73168.1"
                     /db_xref="GI:45644922"
                     /translation="MEKLNFAKVTILSGVVTCFRLICGLLISKVVAIYTGPIGIANLG
                     QLQNFVAFINGFISSQVSQGVNRYSAENQGDYENACFYWRASIKLSFIACCFIISTGV
                     LFSSKISEYLFYQSDLYWLIIIALCALPLNIVNNIFLGVLNGLNEYNRFFLANVLAIS
                     SSLLSMVGLVYFFGLKGALVSASLNNAVAGVWLITIIIKRPWFKFKYWVGHTPRHNIT
                     QMKNYFYMGVIGALTGPISMIVVRTILTNNFSLEDAGYWQAVNRISEAYLAVLTTALT
                     VYYFPKTAAARRYSEYITLLKTGACIVVPLALSMALTIYGLKDFIISILFTADFIRAR
                     ELFLFQNIGDFLRICSWLFATILLAKGYFKINAILEVMFSIIFPVLVKFLIINFGLNA
                     VSIAYSITYALYFIVVCIIYCWHLTKKAL"
     gene            5353. .6339
                     /gene="wbtD"
     CDS             5353. .6339
                     /gene="wbtD"
                     /note="WbtD"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glycosyltransferase"
                     /protein_id="AAS73169.1"
                     /db_xref="GI:45644923"
                     /translation="MIISSDNLSVIIPVYNEEKNVINILQSLECQSLTGFNVIVIDDG
```

```
                SKDMTASLVKEYKPSSYSLSLIQQSNMGAARARENAINFTHSEYIAFIDSDDSLSSDA
                LEKALAPMLDNKDIDISLFELVHIKNLNHDTNNIFVPYSTTKLICGEEAFANCISYWG
                LHGFGIYKRKLIQKSYGIYYKYNKNKENYINNDEVISRISFGLSKNIYLSSGRYFFVQ
                NMDSTTRRINESYYKVINNAVYLREYIDEEIKYNDFDCLGEANKLLVSTIWGVFVRYQ
                KWKKKFSDETNGKWREAIKKGMQYIKKINNNKNINLHIKSKMQLYIISKLVS"
     gene       6342. .7232
                /gene="wbtE"
     CDS        6342. .7232
                /gene="wbtE"
                /note="WbtE"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyltransferase"
                /protein_id="AAS73170.1"
                /db_xref="GI:45644924"
                /translation="MKNLLITIIVASYNSKDYILETLESCINQQYQNIEIIIVDDCSS
                DNSVELIKQWCKEKKSTSPKIRCILVESERNKGIPANLNNALPFINGDWIKCIGSDDI
                LLPEAISEFVNRLEKCSNLDNIGAVFTYFQTFGYSVNVSTRYPSSWTRTICQMPPSWL
                KKAAAMIHFNNVAPCAFINKNHFSGFDTSYKLLEDLPLWLKMIKNNKNTLFFDYTTVL
                YRLHDSQITSLNNIKVNDVLINDLCRLNYYRKNNGYYLAYLHHKFNLYCAVKRMRFYR
                YLKIINPLNLLIVLYERVKK"
     gene       7232. .8380
                /gene="wzy"
     CDS        7232. .8380
                /gene="wzy"
                /note="Wzy"
                /codon_start=1
                /transl_table=11
                /product="putative O-antigen polymerase"
                /protein_id="AAS73171.1"
                /db_xref="GI:45644925"
                /translation="MYILSMAIPLFFISLCWHRFSHQTQVAFRLLFIVFFSLSFFSYY
                TNGVDWSVYYLRFIEDEGLFTSFEFGFVIILKLLLIVSDDNFGLAILLYYFLCFVLLS
                LILKKYKVNEPLFLGCLLLLFGYNLFLEQLRQLLACIIVFYAMLLYNQNKNLIELCIL
                VITASTFHVSAVIILPTIFLISFRNVTTFIIITVSSITSIVVFMIALYSAISSLAAMS
                FVFAKIDYYLHQYPVVLNFGWLNILDALFILFYMRYRNAIDRQININILTRIIFVGSV
                IHLFSGSITFLARVCFYFYFVGVYIYCYSLNDKTPRVFAIKNFNTLVLAAFVNVVLLL
                NFASYFRNEQAPVGFYNMNFQFFNTFNDNYLRTNAYDKYFKGMQGSVD"
     gene       8383. .9474
                /gene="wbtF"
     CDS        8383. .9474
                /gene="wbtF"
                /note="WbtF"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyltransferase"
                /protein_id="AAS73172.1"
                /db_xref="GI:45644926"
                /translation="MSKRIIFVITGLGLGGAEKQICLLADKLYLHNNEVNIVVLNGNV
                NVTPLNENIRIHNLHMKKNPLGLLNAVLRLVKIISYFKPDIVHGHMFHANIVARLAKV
                FTYNKYKLICTAHSKNEGGHIRMLVYRLTDFLCDATTNVSNEALDVFIKKKAFSKSKS
                IPVYNGIDTDLFKYDPLQRDKIRHSLNIINDDSLILSVGRLTDAKDYPNLLKALLELP
                ERYKLVIIGEGEARNEVERTILNYGLEKRVKLLGSIANVATYYSACDIYVSSSKWEGF
                GLVVAEAMSCQRLVVATNAGGVAEVVGDMNYIVPVSDPQLLANKINEVMEFDHETKME
                VMYRNREFVLKMFSIDMIVNQWMKLYSSI"
     gene       9660. .10781
                /gene="wbtG"
     CDS        9660. .10781
                /gene="wbtG"
                /note="WbtG"
                /codon_start=1
                /transl_table=11

  Query Match             22.5%;  Score 252.8;  DB 1;  Length 12033;
  Best Local Similarity   54.3%;  Pred. No. 1.3e-31;
  Matches  593;  Conservative    0;  Mismatches  472;  Indels   27;  Gaps    3;

Qy          7 ATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACGCAGTTA 66
              ||| ||| | | || || |||||    |  ||||| || ||| | |  ||      ||||
Db       8395 ATTATTTTTGTTATTACTGGGCTGGGATTAGGTGGGGCTGAAAAACAAATATGTTTGTTA 8454

Qy         67 GCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGTATATCT 126
              || ||  || ||| | | |      |||| |||||||||| |     | | ||| |

Db       8455 GCTGATAAATTGTATTTGCATAATAATGAAGTTAATATTGTCGTGTTGAATGGAAATGTC 8514

Qy         127 GAGGTAAAGCCA---ACACAAAATATTAATATTCATTATGTTAATATGGATAAAAATTTT 183
               | || |  |||    |   | ||||||  ||||| || |  |||| | ||||||
Db       8515 AATGTTACTCCATTGAATGAGAATATTCGAATTCATAATTTGCATATGAAAAAAAATCCA 8574

Qy         184 AGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCAGAT 243
               | ||  | |  ||  | |      |  |||||||||       || ||||| ||
Db       8575 CTTGGGTTACTAAATGCAGTGCTCAGGCTCGTGAAAATAATTAGTTATTTCAAGCCGGAC 8634

Qy         244 ATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTGATT 303
               ||  | ||| | |||||||| ||||| |||||  ||   |  ||  |  | | |
Db       8635 ATTGTTCATGGACATATGTTCCATGCAAATATCGTTGCCAGACTTGCAAAAGTTTTTACC 8694

Qy         304 CCAGCGGTGCCC---CTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 360
                         |   | || ||||| ||||| |  ||||||||||||||  ||   |||
Db       8695 TATAATAAATACAAATTAATTTGTACAGCACATAGTAAAAATGAAGGTGGGCATATTAGG 8754

Qy         361 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 420
               ||| |   |||| || | | |||||||||| |  |  ||| || ||||||||||| ||
Db       8755 ATGCTCGTTTATAGATTAACTGATTTTTTATGTGATGCTACGACTAATGTAAGTAATGAA 8814

Qy         421 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 480
               ||| || | |  |||||    | ||| ||          |||   ||  | |   ||
Db       8815 GCTCTTGATGTTTTTATTAAGAAAAAAGCCTTTAGCAAATCGAAGTCAATTCCCGTTTAT 8874

Qy         481 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540
               |||   ||| | ||  ||  |||| | | ||||  |       ||| | ||||  |  ||
Db       8875 AATGGCATTGACACTGATCTATTTAAATATGATCCTTTACAAAGAGATAAAATAAGGCAT 8934

Qy         541 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 600
               |||  ||| | |  |    |  |  | | ||  | |||||  |||||  ||| |||
Db       8935 AGTTTAAATATCATTAATGATGATAGCCTTATACTTTCTGTAGGCCGACTTACTGATGCA 8994

Qy         601 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 660
               |||||  ||||| ||||||||||| ||| |  |  ||   |  | | | || |
Db       8995 AAAGATTATCCAAACTTATTAAAAGCATTGTTAGAATTACCTGAACGATATAAATTAG-- 9052

Qy         661 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 720
                                 | ||    || || || | |   |||||| || |   | |
Db       9053 ------------------TCATAATAGGTGAAGGTGAAGCCAGAAATGAAGTCGAGCGT 9093

Qy         721 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 780
               | |||   | |    ||||  | | |||   ||  | |||      |   | || |
Db       9094 ACGATTTTAAATTACGGCCTTGAAAAAAGAGTGAAATTGCTAGGTAGTATTGCTAATGTC 9153

Qy         781 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 840
                  |    | |||    ||| ||| ||| | |   ||| ||||||||| ||||||
Db       9154 GCCACTTATTATTCTGCATGTGATATTTATGTATCGTCCTCTAAGTGGGAAGGCTTTGGT 9213

Qy         841 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 900
               | ||||| || || |||||| | ||  | ||   || ||||| ||  || | |||||
Db       9214 TTGGTTGTAGCTGAGGCTATGTCTTGCCAGCGATTGGTCGTTGCAACTAATGCCGGTGGG 9273

Qy         901 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 960
               ||    ||||| |||||    |   ||  || || | |||||      | | |   |||||
Db       9274 GTGGCTGAAGTTGTTGGTGATATGAACTATATTGTACCTGTTTCAGACCCTCAGTTGTTG 9333

Qy         961 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 1020
               ||| | |||||   |||      |  |  | ||| |  |  | | ||    | ||
Db       9334 GCAAATAAAATTAATGAAGTTATGGAGTTTGATCACGAAACTAAAATGGAAGTAATGTAT 9393

Qy        1021 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 1080
               | |||||||| | | || |     | || |||||  | | ||| || | | | ||| |
Db       9394 AGAAATAGAGAGTTTGTTTTAAAAATGTTCTCAATAGATATGATCGTTAATCAATGGATG 9453

Qy        1081 CGCTTATATTTT 1092
                  |||| |
Db       9454 AAATTGTATTCT 9465


RESULT 7
AX276712
```

```
LOCUS       AX276712               1104 bp    DNA     linear   PAT 29-OCT-2001
DEFINITION  Sequence 1321 from Patent WO0166572.
ACCESSION   AX276712
VERSION     AX276712.1  GI:16548511
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Bingen,E., Bonacorsi,S., Clermont,O., Nassif,X. and Tinsley,C.
  TITLE     Polynucleotides which are of nature b2/d+ a- and which are isolated
            from E. Coli, and biological uses of these polynucleotides and of
            their polypeptides
  JOURNAL   Patent: WO 0166572-A 1321 13-SEP-2001;
            INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM)
            (FR)
FEATURES             Location/Qualifiers
     source          1. .1104
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
ORIGIN

  Query Match             21.8%;  Score 245.8;  DB 6;  Length 1104;
  Best Local Similarity   55.8%;  Pred. No. 3.6e-30;
  Matches  550;  Conservative    0;  Mismatches  402;  Indels   33;  Gaps    3;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||| ||  | | | | || || ||  |      ||| || || ||  | |   |
Db          1 ATGAAAATACTATTTGTCATTACAGGTTTAGGCCTTGGAGGTGCTGAGAAGCAGGTTTGT 60

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
              |  |||||| ||  || |   | || |||||||  |||| || |||||||  ||      |
Db         61 CTTTTAGCTGATAAATTAAGTTTAAGCGGGCACCATGTAAAGATTATTTCACTTGGACAT 120

Qy        121 ATATCTGAGGTAAAGCCAACAC---------AAAATATTAATATTCATTATGTTAATATG 171
              || ||| |    ||        |         | ||| ||||| |   | |||| ||||||
Db        121 ATGTCTAATAATAAAGTCTTTCCTAGCGAAAATAATGTTAATGTCATTAATGTAAATATG 180

Qy        172 GATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCC 231
              |||||  ||    |  || | | || ||   |   | | | | |  | || |
Db        181 TCAAAAAACATTTCTGGAGTTATAAAAGGTTGTGTCAGAATTAGAGATGTTATAGCTAAT 240

Qy        232 TTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATT 291
              || || ||||| ||  |||| |||||||||||||||| || ||| | | | | ||   |
Db        241 TTCAAACCAGACATTGTACACAGTCATATGTTTCATGCAAACATTATCACTAGATTGTCT 300

Qy        292 AGGATGCTGATTCCAGCGGTGCC---CCTGATATGTACCGCACACAACAAAAATGAAGGT 348
              ||   ||   |    ||     | ||||  || ||||| || ||||||||||||
Db        301 GTAATTGGAATCAAAAACAGACCTGGTATTATATCAACTGCACATAATAAAAATGAAGGT 360

Qy        349 GGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAAT 408
              ||  ||   || ||| |    ||| || | |  |||| ||||  |  |  |  |||||||
Db        361 GGGTATTTCAGAATGCTCACATATAGAATAACCGATTGTTTAAGTGATTGTTGTACAAAT 420

Qy        409 GTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATA 468
              || || ||||| || ||  | |||||| ||   | ||| ||       |     |||| |
Db        421 GTTAGCAAAGAAGCAGTGGATGAGTTTTTACGGATAAAAGCCTTTAATCCCGCTAAAGCA 480

Qy        469 GTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAG 528
              |    ||    |||   ||  |||| |||||||| | ||||||  |     |  || | |
Db        481 ATTACTATGTATAATGGGATAGATACCAATAAATTTAAATTTGATTTATTGGCAAGGAGG 540

Qy        529 AAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGA 588
              |||  ||||| | | ||||| | ||| |    |     | || | || || | |||  |
Db        541 GAAATTCGAGACGGTATTAATATAAAAAATGATGATATATTATTACTTGCTGCAGGTCGT 600

Qy        589 CTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCA 648
              |      ||| ||||| ||||| || ||||| |||||||| | ||   |
Db        601 TTAACGTTAGCTAAAGATTATCCTAATTTATTGAATGCAATGACTCTGCT---------- 650

Qy        649 AAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAAT 708
```

```
                  |  | |  | ||||   |  |||||  ||| ||||| | |||  |  |
Db        651 ----------TCCTGAACACTTTAAACTTATTATTATTGGTGATGGTGAATTGCGTGAC 699

Qy        709 AAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAA 768
              || |      |  |     ||||| |  |     |||  || | ||| |||||
Db        700 GAAATTAATATGCTTATAAAAAAAATTGCAATTATCTAATAGGGTGTCCTTGTTGGGAGTT 759

Qy        769 AGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGG 828
              | ||   |||||       |   | | |||    ||| ||||||||  |   ||||   |||
Db        760 AAAAAAAATATTGCTCCCTATTTTTCTGCATGTGATATTTTTGTTCTCTCTTCTCGTTGG 819

Qy        829 GAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACC 888
              |||||  |||||  | || || ||||||||||||| | |||||  ||    |||||||  ||
Db        820 GAAGGATTTGGATTAGTCGTGGCAGAAGCTATGTCATGTGAGCGAATTGTTGTTGGCACG 879

Qy        889 GATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAAT 948
              ||||| || ||||| | ||||||  ||||     ||  ||  | ||  | ||   |  | ||
Db        880 GATTCAGGGGGAGTAAGAGAAGTTATTGGTGACGATGATTTTCTTGTACCCATATCTGAT 939

Qy        949 CATATTCTGTTGGCAGAGAAAATCG 973
              |  |   | |||   ||||| |
Db        940 TCAACACAACTTGCAAGCAAAATTG 964
```

```
RESULT 8
BD248530
LOCUS       BD248530               4604 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Virulence genes and proteins, and their use.
ACCESSION   BD248530
VERSION     BD248530.1  GI:33058300
KEYWORDS    JP 2002529091-A/10.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 4604)
  AUTHORS   Crooke,H.R., Clarke,E.E., Everest,P.H., Dougan,G., Holden,D.W.,
            Shea,J.E. and Feldman,R.G.
  TITLE     Virulence genes and proteins, and their use
  JOURNAL   Patent: JP 2002529091-A 10 10-SEP-2002;
            MICROSCIENCE LTD
COMMENT     OS   Escherichia coli
            PN   JP 2002529091-A/10
            PD   10-SEP-2002
            PF   09-NOV-1999 JP 2000581205
            PR   09-NOV-1998 GB    9824569.9,09-NOV-1998 GB    9824570.7 PR
            17-DEC-1998 GB    9827814.6,17-DEC-1998 GB    9827815.3 PR
            17-DEC-1998 GB    9827816.1,17-DEC-1998 GB    9827818.7 PR
            13-JAN-1999 GB    9900708.0,13-JAN-1999 GB    9900710.6 PR
            13-JAN-1999 GB    9900711.4,28-JAN-1999 GB    9901915.0 PI   HELEN
            RACHEL CROOKE,ENDA ELIZABETH CLARKE,PAUL HOWARD EVEREST, PI
            GORDON DOUGAN,DAVID WILLIAM HOLDEN,JACQUELINE ELIZABETH SHEA, PI
            ROBERT GRAHAM FELDMAN
            PC   C12N15/09,A61K35/74,A61K38/00,A61K39/108,A61K48/00,A61P31/04,
            PC   C07K14/345,
            PC   C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/02,C12Q1/02,C12Q1/
            PC   68,G01N33/15,
            PC   G01N33/50,C12N15/00,C12N5/00,A61K37/02
            CC   Virulence genes and proteins, and their use
            FH   Key             Location/Qualifiers
            FT   CDS             (2982). .(4082)
            FT   CDS             (1534). .(2637)
            FT   CDS             (749). .(1531)
            FT   CDS             (6). .(746).
FEATURES             Location/Qualifiers
     source          1. .4604
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:562"
ORIGIN

  Query Match             21.8%;  Score 245.8;  DB 6;  Length 4604;
  Best Local Similarity   55.8%;  Pred. No. 2.4e-30;
```

```
  Matches  550;  Conservative    0;  Mismatches  402;  Indels   33;  Gaps    3;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
Db       1534 ATGAAAATACTATTTGTCATTACAGGTTTAGGCCTTGGAGGTGCTGAGAAGCAGGTTTGT 1593

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
Db       1594 CTTTTAGCTGATAAATTAAGTTTAAGCGGGCACCATGTAAAGATTATTTCACTTGGACAT 1653

Qy        121 ATATCTGAGGTAAAGCCAACAC---------AAAATATTAATATTCATTATGTTAATATG 171
Db       1654 ATGTCTAATAATAAAGTCTTTCCTAGCGAAAATAATGTTAATGTCATTAATGTAAATATG 1713

Qy        172 GATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCC 231
Db       1714 TCAAAAAACATTTCTGGAGTTATAAAAGGTTGTGTCAGAATTAGAGATGTTATAGCTAAT 1773

Qy        232 TTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATT 291
Db       1774 TTCAAACCAGACATTGTACACAGTCATATGTTTCATGCAAACATTATCACTAGATTGTCT 1833

Qy        292 AGGATGCTGATTCCAGCGGTGCC---CCTGATATGTACCGCACACAACAAAAATGAAGGT 348
Db       1834 GTAATTGGAATCAAAAACAGACCTGGTATTATATCAACTGCACATAATAAAAATGAAGGT 1893

Qy        349 GGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAAT 408
Db       1894 GGGTATTTCAGAATGCTCACATATAGAATAACCGATTGTTTAAGTGATTGTTGTACAAAT 1953

Qy        409 GTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATA 468
Db       1954 GTTAGCAAAGAAGCAGTGGATGAGTTTTTACGGATAAAAGCCTTTAATCCCGCTAAAGCA 2013

Qy        469 GTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAG 528
Db       2014 ATTACTATGTATAATGGGATAGATACCAATAAATTTAAATTTGATTTATTGGCAAGGAGG 2073

Qy        529 AAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGA 588
Db       2074 GAAATTCGAGACGGTATTAATATAAAAAATGATGATATATTATTACTTGCTGCAGGTCGT 2133

Qy        589 CTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCA 648
Db       2134 TTAACGTTAGCTAAAGATTATCCTAATTTATTGAATGCAATGACTCTGCT---------- 2183

Qy        649 AAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAAT 708
Db       2184 ----------TCCTGAACACTTTAAACTTATTATTATTGGTGATGGTGAATTGCGTGAC 2232

Qy        709 AAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAA 768
Db       2233 GAAATTAATATGCTTATAAAAAAATTGCAATTATCTAATAGGGTGTCCTTGTTGGGAGTT 2292

Qy        769 AGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGG 828
Db       2293 AAAAAAAATATTGCTCCCTATTTTTCTGCATGTGATATTTTTGTTCTCTCTTCTCGTTGG 2352

Qy        829 GAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACC 888
Db       2353 GAAGGATTTGGATTAGTCGTGGCAGAAGCTATGTCATGTGAGCGAATTGTTGTTGGCACG 2412

Qy        889 GATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAAT 948
Db       2413 GATTCAGGGGGAGTAAGAGAAGTTATTGGTGACGATGATTTTCTTGTACCCATATCTGAT 2472

Qy        949 CATATTCTGTTGGCAGAGAAAATCG 973
Db       2473 TCAACACAACTTGCAAGCAAAATTG 2497

RESULT 9
AR632878

LOCUS       AR632878                4604 bp    DNA     linear   PAT 14-FEB-2005
DEFINITION  Sequence 22 from patent US 6846667.
ACCESSION   AR632878
VERSION     AR632878.1  GI:59778814
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4604)
  AUTHORS   Crooke,H.R., Clarke,E.E., Everest,P.H., Dougan,G., Holden,D.W.,
            Shea,J.E. and Feldman,R.G.
  TITLE     Virulence genes and proteins, and their use
  JOURNAL   Patent: US 6846667-A 22 25-JAN-2005;
            Microscience, Ltd.;;
            GBX;
FEATURES             Location/Qualifiers
     source          1. .4604
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             21.8%;  Score 245.8;  DB 6;  Length 4604;
  Best Local Similarity   55.8%;  Pred. No. 2.4e-30;
  Matches  550;  Conservative    0;  Mismatches  402;  Indels   33;  Gaps    3;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
Db       1534 ATGAAAATACTATTTGTCATTACAGGTTTAGGCCTTGGAGGTGCTGAGAAGCAGGTTTGT 1593

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
Db       1594 CTTTTAGCTGATAAATTAAGTTTAAGCGGGCACCATGTAAAGATTATTTCACTTGGACAT 1653

Qy        121 ATATCTGAGGTAAAGCCAACAC---------AAAATATTAATATTCATTATGTTAATATG 171
Db       1654 ATGTCTAATAATAAAGTCTTTCCTAGCGAAAATAATGTTAATGTCATTAATGTAAATATG 1713

Qy        172 GATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCC 231
Db       1714 TCAAAAAACATTTCTGGAGTTATAAAAGGTTGTGTCAGAATTAGAGATGTTATAGCTAAT 1773

Qy        232 TTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATT 291
Db       1774 TTCAAACCAGACATTGTACACAGTCATATGTTTCATGCAAACATTATCACTAGATTGTCT 1833

Qy        292 AGGATGCTGATTCCAGCGGTGCC---CCTGATATGTACCGCACACAACAAAAATGAAGGT 348
Db       1834 GTAATTGGAATCAAAAACAGACCTGGTATTATATCAACTGCACATAATAAAAATGAAGGT 1893

Qy        349 GGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAAT 408
Db       1894 GGGTATTTCAGAATGCTCACATATAGAATAACCGATTGTTTAAGTGATTGTTGTACAAAT 1953

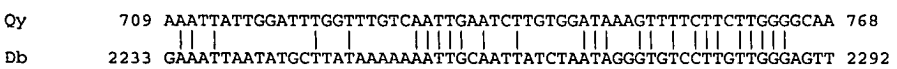
Qy        409 GTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATA 468
Db       1954 GTTAGCAAAGAAGCAGTGGATGAGTTTTTACGGATAAAAGCCTTTAATCCCGCTAAAGCA 2013

Qy        469 GTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAG 528
Db       2014 ATTACTATGTATAATGGGATAGATACCAATAAATTTAAATTTGATTTATTGGCAAGGAGG 2073

Qy        529 AAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGA 588
Db       2074 GAAATTCGAGACGGTATTAATATAAAAAATGATGATATATTATTACTTGCTGCAGGTCGT 2133

Qy        589 CTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCA 648
Db       2134 TTAACGTTAGCTAAAGATTATCCTAATTTATTGAATGCAATGACTCTGCT---------- 2183

Qy        649 AAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAAT 708
Db       2184 ----------TCCTGAACACTTTAAACTTATTATTATTGGTGATGGTGAATTGCGTGAC 2232
```



```
Qy        709 AAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAA 768
Db       2233 GAAATTAATATGCTTATAAAAAAATTGCAATTATCTAATAGGGTGTCCTTGTTGGGAGTT 2292

Qy        769 AGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGG 828
Db       2293 AAAAAAAATATTGCTCCCTATTTTTCTGCATGTGATATTTTTGTTCTCTCTTCTCGTTGG 2352

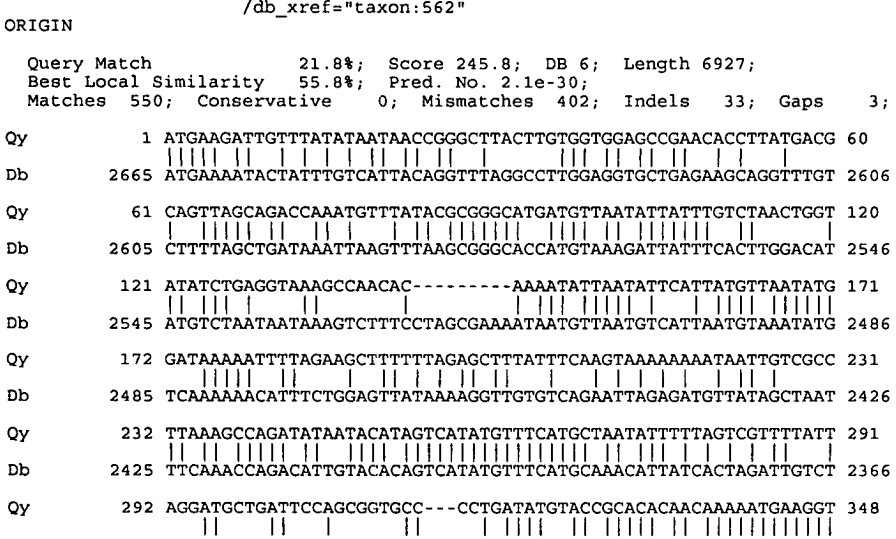
Qy        829 GAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACC 888
Db       2353 GAAGGATTTGGATTAGTCGTGGCAGAAGCTATGTCATGTGAGCGAATTGTTGTTGGCACG 2412

Qy        889 GATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAAT 948
Db       2413 GATTCAGGGGGAGTAAGAGAAGTTATTGGTGACGATGATTTTCTTGTACCCATATCTGAT 2472

Qy        949 CATATTCTGTTGGCAGAGAAAATCG 973
Db       2473 TCAACACAACTTGCAAGCAAAATTG 2497
```

RESULT 10
AX276709/c
LOCUS AX276709 6927 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 1318 from Patent WO0166572.
ACCESSION AX276709
VERSION AX276709.1 GI:16548509
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Bingen,E., Bonacorsi,S., Clermont,O., Nassif,X. and Tinsley,C.
TITLE Polynucleotides which are of nature b2/d+ a- and which are isolated from E. Coli, and biological uses of these polynucleotides and of their polypeptides
JOURNAL Patent: WO 0166572-A 1318 13-SEP-2001;
INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM) (FR)
FEATURES Location/Qualifiers
source 1. .6927
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
ORIGIN

Query Match 21.8%; Score 245.8; DB 6; Length 6927;
Best Local Similarity 55.8%; Pred. No. 2.1e-30;
Matches 550; Conservative 0; Mismatches 402; Indels 33; Gaps 3;



```
Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
Db       2665 ATGAAAATACTATTTGTCATTACAGGTTTAGGCCTTGGAGGTGCTGAGAAGCAGGTTTGT 2606

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
Db       2605 CTTTTAGCTGATAAATTAAGTTTAAGCGGGCACCATGTAAAGATTATTTCACTTGGACAT 2546

Qy        121 ATATCTGAGGTAAAGCCAACAC---------AAAATATTAATATTCATTATGTTAATATG 171
Db       2545 ATGTCTAATAATAAAGTCTTTCCTAGCGAAAATAATGTTAATGTCATTAATGTAAATATG 2486

Qy        172 GATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCC 231
Db       2485 TCAAAAAACATTTCTGGAGTTATAAAAGGTTGTGTCAGAATTAGAGATGTTATAGCTAAT 2426

Qy        232 TTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATT 291
Db       2425 TTCAAACCAGACATTGTACACAGTCATATGTTTCATGCAAACATTATCACTAGATTGTCT 2366

Qy        292 AGGATGCTGATTCCAGCGGTGCC---CCTGATATGTACCGCACACAACAAAAATGAAGGT 348
Db       2365 GTAATTGGAATCAAAAACAGACCTGGTATTATATCAACTGCACATAATAAAAATGAAGGT 2306

Qy        349 GGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAAT 408
Db       2305 GGGTATTTCAGAATGCTCACATATAGAATAACCGATTGTTTAAGTGATTGTTGTACAAAT 2246

Qy        409 GTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATA 468
Db       2245 GTTAGCAAAGAAGCAGTGGATGAGTTTTTACGGATAAAAGCCTTTAATCCCGCTAAAGCA 2186

Qy        469 GTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAG 528
Db       2185 ATTACTATGTATAATGGGATAGATACCAATAAATTTAAATTTGATTTATTGGCAAGGAGG 2126

Qy        529 AAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGA 588
Db       2125 GAAATTCGAGACGGTATTAATATAAAAAATGATGATATATTATTACTTGCTGCAGGTCGT 2066

Qy        589 CTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCA 648
Db       2065 TTAACGTTAGCTAAAGATTATCCTAATTTATTGAATGCAATGACTCTGCT---------- 2016

Qy        649 AAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAAT 708
Db       2015 ----------TCCTGAACACTTTAAACTTATTATTATTGGTGATGGTGAATTGCGTGAC 1967

Qy        709 AAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAA 768
Db       1966 GAAATTAATATGCTTATAAAAAAATTGCAATTATCTAATAGGGTGTCCTTGTTGGGAGTT 1907

Qy        769 AGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGG 828
Db       1906 AAAAAAAATATTGCTCCCTATTTTTCTGCATGTGATATTTTTGTTCTCTCTTCTCGTTGG 1847

Qy        829 GAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACC 888
Db       1846 GAAGGATTTGGATTAGTCGTGGCAGAAGCTATGTCATGTGAGCGAATTGTTGTTGGCACG 1787

Qy        889 GATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAAT 948
Db       1786 GATTCAGGGGGAGTAAGAGAAGTTATTGGTGACGATGATTTTCTTGTACCCATATCTGAT 1727

Qy        949 CATATTCTGTTGGCAGAGAAAATCG 973
Db       1726 TCAACACAACTTGCAAGCAAAATTG 1702
```

RESULT 11
AY057452
LOCUS AY057452 12215 bp DNA linear BCT 03-JUN-2003
DEFINITION Edwardsiella ictaluri isolate 93-146 O antigen biosynthesis gene cluster, partial sequence; and IS1-like insertion sequence, complete sequence.
ACCESSION AY057452
VERSION AY057452.2 GI:19113666
KEYWORDS .
SOURCE Edwardsiella ictaluri
ORGANISM Edwardsiella ictaluri
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Edwardsiella.
REFERENCE 1 (bases 1 to 12215)
AUTHORS Lawrence,M.L., Banes,M.M., Azadi,P. and Reeks,B.Y.
TITLE The Edwardsiella ictaluri O polysaccharide biosynthesis gene cluster and the role of O polysaccharide in resistance to normal catfish serum and catfish neutrophils
JOURNAL Microbiology (Reading, Engl.) 149 (Pt 6), 1409-1421 (2003)
PUBMED 12777482
REFERENCE 2 (bases 1 to 12215)
AUTHORS Lawrence,M.L. and Banes,M.M.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-2001) Veterinary Basic Sciences, Mississippi State University, P.O. Box 6100, Mississippi State, MS 39762-6100, USA

```
REFERENCE   3  (bases 1 to 12215)
  AUTHORS   Lawrence,M.L. and Banes,M.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2002) Veterinary Basic Sciences, Mississippi
            State University, P.O. Box 6100, Mississippi State, MS 39762-6100,
            USA
  REMARK    Sequence update by submitter
COMMENT     On Mar 5, 2002 this sequence version replaced gi:16648654.
FEATURES             Location/Qualifiers
     source          1. .12215
                     /organism="Edwardsiella ictaluri"
                     /mol_type="genomic DNA"
                     /isolate="93-146"
                     /db_xref="taxon:67780"
     gene            <1. .1204
                     /gene="dcuC"
     CDS             <1. .1204
                     /gene="dcuC"
                     /codon_start=2
                     /transl_table=11
                     /product="C4-dicarboxylate anaerobic carrier"
                     /protein_id="AAL25626.1"
                     /db_xref="GI:16648655"
                     /translation="STGWRATDIVEYVKVLLMSRGGDLGMMIMMLCGFAAYMTHIGAN
                     DMVVKLASRPLRYINSPYILMIAAYFVACLMSLAVSSATGLGVLLMATLFPVMVNVGI
                     SRGAAAAICASPAAIILAPTSGDVVLAAKASEMPLIDFAFKTTLPISIIAIVCMAIAH
                     YFWQRYLDNKEQAQHEILDVNEITTTAPSFYAILPFTPIIGVLVFDGKWGPHLHIISV
                     LVICMLLASVTEFLRSFDAKKVYSGLEVAYRGMADAFASVVMLLVAAGVFAQGLSTVG
                     FISSLIELAQSFGTGGIIMMLVLVVITMLAAMTTGSGNAPFYAFVELIPKLSGQMGIN
                     PAYLTIPMLQASNLGRTLSPVSGVVVAVSGMAKISPFEVVKRTSVPVLVGLVVVIVAT
                     EFMVPLHR"
     repeat_region   2079. .2746
                     /insertion_seq="IS1-like"
     gene            2079. .2312
                     /gene="insA"
                     /note="similar to InsA protein"
     gene            2354. .2746
                     /gene="insB"
                     /note="similar to InsB protein"
     gene            2984. .4231
                     /gene="wzx"
     CDS             2984. .4231
                     /gene="wzx"
                     /codon_start=1
                     /transl_table=11
                     /product="O antigen transporter"
                     /protein_id="AAL25627.1"
                     /db_xref="GI:16648656"
                     /translation="MSSLLKNSISLLAMQGINYILPLVTLPYLTRVLHVYPFGIYSAT
                     MSVGAYVILFINFGFDLSATRKIAKNKGDNYAISAIFFNVILAKFLLFFLSLVIVFLL
                     LYTIDSFYPIRSLFFLLIPQFIGCVFFPAWLYQGVERLYLISAITSATKLLIIPLFFL
                     FVRDAGDIDKAIIISGFPLFFSGIIGFYLAYKLNIINFSDVRVRVNDAILLIKESTSV
                     FIGAVAISIYTACTPIILSIVSNYTEVGYFSAVDKLRSAILGMFIVLGQVFYPRATVL
                     FESNFSKYKLFIRRIIFSQIIFGGVGTILFYFFMPIIAPFILGRGFHGLTELIEIMAP
                     MILLIPLSVIFSNCILLPMGKSKLFYIAPIITAIIHLPYSIYLSEKYGAFGGGVSILI
                     TEVLSLMLLVFFSYKYANLKSIF"
     gene            4791. .5954
                     /gene="wbeiA"
     CDS             4791. .5954
                     /gene="wbeiA"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="WbeiA"
                     /protein_id="AAL25628.1"
                     /db_xref="GI:16648657"
                     /translation="MFPTKEFKYFDFFCVIEGLFKVLLSKLSVNRSVNLLFFYLFFSV
                     FVLIFIFKPPFHFFLLNALFIHVYSLVALIFSFFYRVKRKYIIIFFIFLFPLSYLTIM
                     DRVTAAYPLVVALFSLLIISDKFQFHNYSCRNLTHFILISTVLFSFLFLYESHIDRYS
                     LVNGDPNYTSLILLTVLLISFVFSVSNKLKVLVLLLMLFVVYLTKSRTAILALAAFFL
                     CCKIKRKLLLFYFILFVSIASQYFFSYIFPSILSDLDFGHSRFFDLNDASNMERITIY
                     KEAVDFILSHFTHFIINGVTNYLEINKLATNIPHNWFIQFILGFGFLFSIYYLAVVVF
                     VAKIVSGKSYYLLAFMCFFFVYASFLSYYSLATPLLFLVLSLFVCHSIERVAR"

     gene            6029. .6814
                     /gene="wbeiB"
     CDS             6029. .6814
                     /gene="wbeiB"
                     /note="includes previous wbeiC gene, whose presence was
                     annotated because of an error in the sequence"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glycosyltransferase"
                     /protein_id="AAL25629.2"
                     /db_xref="GI:19113667"
                     /translation="MKTFIISLKDEVARRNSISDRLHHHDFEFFDAVDLRVASDDVLS
                     AVKCKNINYKHPAIREHMTKGEIGCALSHMQLYKKIVDDNLAFARVIEDDAVILNHDE
                     KILNDFILALEMKNIDWDIMLLGYSKLKACDSFGFYLKEPIKNIVKSGAYSLGIPFRN
                     WTCGTVGYLVSQSGAKKMINNAAFGKVCTVADDWLFFEKNYNLKILHIRPLIVLEDFE
                     SYESSIETDRKTYVNHESHISFFLRVMRGVVRKVYLSISWKSL"
     gene            6811. .7911
                     /gene="wbeiM"
     CDS             6811. .7911
                     /gene="wbeiM"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glycosyltransferase"
                     /protein_id="AAL25631.1"
                     /db_xref="GI:16648660"
                     /translation="MKILYVITGLGLGGAEKQLSLLADNFTARGEQVSIVYLTGEVLV
                     KPKNKNIKIYNLGIDKSFSSLIKGIWKLKSIISDVRPDVIHSHMYHANILARISCCLS
                     LFSSRLVCSAHNKNEGGRVRMIIYRMTDFLCAKTTNVSQEALDEFITKKAFRKRKSSL
                     VYNGIDLSIFKKKSTNIQNIKNKLGINFDEKVIFCAGRLTEAKDYPNLILAISKMHQK
                     KCKIIIAGDGPMRSDIERLIDRCHLSHRILLIGIIDNISDYYNLSDLFVLPSRWEGFG
                     LVVAEAMACECPVIATDAGGVAEVLSNADWLVPIADSSKLAEKIDEFFLLDSSEVKDI
                     KAKNKDHCENQFSIGAIINNWYAIYNGIRSKL"
     gene            7913. .8950
                     /gene="wbeiH"
     CDS             7913. .8950
                     /gene="wbeiH"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glycosyltransferase"
                     /protein_id="AAL25632.1"
                     /db_xref="GI:16648661"
                     /translation="MKKIPFVITKSEIGGAQKWIVELSNILKEQYEIFIITSGDGWLT
                     SKFDSSHVILYPGLLNLKSIFCFFKLANIFRKYKADIVISSSANAGVYARLSKFLYSH
                     KSIYVSHGWSCLYNGGRLKKIQCLVERFLSLISDKILCVSENDMNNAIKFIGIKEKKL
                     FVIRNGVSPLKKKNVINVRKKVLFVGRMIHPKRPDLFVAVADSFPDVDFYLVGDGPLK
                     QRLETSCQHNNVYFLGEINGFDSFQCYDIFVLCSDSEGLPMSALEAASSGVPILLSDV
                     GGCSELITLMNNGCSNGELFSNDISSLRCKLGAVLLEYDKYFSSAQVVSGCFDINNLA
                     EKYIELIEGDK"
     gene            8947. .9918
                     /gene="wbeiT"
     CDS             8947. .9918
                     /gene="wbeiT"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-glucose 4-epimerase"
                     /protein_id="AAL25633.1"
                     /db_xref="GI:16648662"
                     /translation="MKLAIIGGSGFIGTRLATQLKNNNVEFTIVDKRNSETFPECWQY
                     GDVTKPESLEEPLKGADVIINLAAEHKDNVHPISLYYDVNVQGAKNVCDVAEKLNIQH
                     IVFTSSVAVYGFVEKETGEDGKYNPFNDYGKSKLEAEFAYDAWQAKVPERTLVTVRPT
                     VVFGENNRGNVYNLFRQIASGKFLMIGPGLNQKSMAYVENMAAFLRFSTTMGAGRHVF
                     NYIDKPDFSMNDLTGVICTALHRHKNNIRIPYAVGLLGGYCFDILAKVTGREFPVSSI
                     RVKKFCARTQFKSNHIAQTGFKAPVTLEQGIANTVRYEFLEKHDDKS"
     gene            9905. .10807
                     /gene="galF"
     CDS             9905. .10807
                     /gene="galF"
                     /codon_start=1
                     /transl_table=11
                     /product="UTP-glucose-1-phosphate uridylyltransferase"
                     /protein_id="AAL25634.1"
                     /db_xref="GI:16648663"
                     /translation="MTSLKAVIPVAGLGMKMLPATKAIPKEMLPIVDKPLIQKVVHEC
```

```
                     LSAGVKEIILVTHSSKNTIENHFDTSFELEALLEARVKRQLLDEVRSICPPGVTINSV
                     RQHRQLGLGHAILCARPLVGEHPFIVALPDIWLDASTFDPETQNLAAMVRRFEESGRS
                     QVLVQPMPLEALPEYSVINCADAVLEPGQSACMTSIVEKPEDAEKYHSDLSAVGRYVL
                     SAAIWPILAATGFGAWERIQLSDAIATLLQHRPVEAYRQVGLSYDCGDKLGYAEAFVA
                     GGLNHSQQGAAFTAWLQDVQTRIE"
     gene            10823. .11839
                     /gene="galE"
     CDS             10823. .11839
                     /gene="galE"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-galactose 4-epimerase"
                     /protein_id="AAL25635.1"
                     /db_xref="GI:16648664"
                     /translation="MAILVTGGAGYIGSHTVLALLSRGDEVVVLDNFVNASVQALARV
                     EQICDRQPHLYRGDVRDRALLQQIFAQHVISDVIHFAGLKAVGESVEKPLEYYDNNVS
                     GTLVLLEEMRRAGVQGFIFSSSATVYGDPESVPLTESSRTGGTTNPYGTSKLMVEQIL
                     QDVSRARPEMRITALRYFNPVGAHPSGLIGEDPNGIPNNLIPYVAQVAVGKLACLSIF
                     GHDYPTQDGTGVRDYIHVMDLATGHLAALDRRDEGASYKVYNLGTGVGYSVLDIVAAF
                     ERCSGVTIPYRFAPRRAGDIAECWSNPALAAQELGWCAQYGLEAMMTDTWRWQQNNPN
                     GYQD"
     gene            11863. .>12215
                     /gene="ugd"
     CDS             11863. .>12215
                     /gene="ugd"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-glucose 6-dehydrogenase"
                     /protein_id="AAL25636.1"
                     /db_xref="GI:16648665"

  Query Match             20.4%;  Score 229.8;  DB 1;  Length 12215;
  Best Local Similarity   52.9%;  Pred. No. 6.9e-28;
  Matches  579;  Conservative    0;  Mismatches  492;  Indels   24;  Gaps    3;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||| ||  | ||| | | || ||| |    | ||||| || ||  | |   |  |
Db       6811 ATGAAAATACTATATGTTATTACAGGGTTAGGGTTGGGTGGGGCAGAGAAACAACTTTCT 6870

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
                |||||||  |  |   |    | ||  |  | |||   ||  ||| | |||| ||
Db       6871 TTATTAGCAGATAATTTTACTGCTAGAGGTGAGCAAGTTTCAATCGTTTATTTAACCGGA 6930

Qy        121 ATATCTGAGGTAAAGCCAACAC---AAAATATTAATATTCATTATGTTAATATGGATAAA 177
                 ||| ||||| | |    ||||||| || ||  || || |   ||| ||||||
Db       6931 GAGGTACTGGTTAAGCCGAAAAACAAAAATATAAAAATATATAATCTAGGTATAGATAAA 6990

Qy        178 AATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAG 237
              | ||||       |  | | |   | |   || |||||  ||||||    | | | | |
Db       6991 AGTTTTTCTTCACTAATCAAGGGGATCTGGAAACTAAAATCAATAATATCTGACGTTAGG 7050

Qy        238 CCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATG 297
              || ||| | || |||   ||||||| |||||||||||| ||   |||| ||   |
Db       7051 CCTGATGTGATCCATTCGCATATGTATCATGCTAATATCTTAGCTCGTATTTCATGTTGC 7110

Qy        298 CTGATTCCAGCGGTGCCCCTGATA---TGTACCGCACACAACAAAAATGAAGGTGGCAAT 354
              ||    | |       |   | ||   ||  | || || || ||||| || || ||
Db       7111 CTTTCCCTATTTTCATCTAGGTTAGTGTGCTCGGCTCATAATAAAAACGAGGGGGGGCGA 7170

Qy        355 GCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGT 414
              |  |||||| ||  ||||||| ||| ||||||| ||  | |||  || |||||||| |||
Db       7171 GTTAGGATGATTATTTATCGAATGACTGATTTTCTATGTGCTAAAACAACAAATGTTAGT 7230

Qy        415 AAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAG 474
              | || ||| |  | |||||||||| | |  || ||       ||||  ||
Db       7231 CAGGAAGCTCTGGATGAGTTTATAACTAAGAAAGCATTTAGAAAAAGAAAGTCTAGTTTA 7290

Qy        475 ATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACG 534
              |   |||  ||  ||     ||  ||| |     | |  | |   |  | || |
Db       7291 GTCTATAATGGAATAGATCTTTCTATCTTTAAAAAGAAAAGCACTAATATTCAAAATATT 7350

Qy        535 CGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTT 594
                | ||   ||     | ||      |      ||  |  |    | ||| ||| |

Db       7351 AAAAATAAATTAGGAATAAATTTTGATGAGAAGGTGATATTTTGTGCAGGTAGATTAACA 7410

Qy        595 GAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACA 654
              |||||||||||||||| || ||  |     ||||||  | |  ||   |    ||||
Db       7411 GAAGCAAAAGACTATCCTAATTTGATCCTCGCAATATCTAAGATGCACCAGAAAAAA--- 7467

Qy        655 TCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTA 714
                          | ||   |  | |||||||| ||||  |  |  | | | |  ||
Db       7468 --------------TGTAAGATTATCATTGCTGGTGATGGTCCCATGCGTAGTGATATA 7512

Qy        715 TTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGT 774
                |    |  ||  ||  |   ||||   | ||    ||||  |  | |||  ||    |
Db       7513 GAGAGACTAATTGATCGTTGCCATCTATCGCATCGGATTTTACTTATTGGGATAATTGAT 7572

Qy        775 GATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGT 834
              |||||   || |        |     | ||||| ||||| |||   |||   ||||| |||
Db       7573 AATATTTCTGATTACTATAATTTATCTGATCTATTTGTATTGCCATCTCGCTGGGAGGGT 7632

Qy        835 TTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCT 894
              |||||  | ||||| ||||| |||||||| |||||| | || ||  |  || ||| |
Db       7633 TTTGGGTTAGTTGTCGCAGAGGCTATGGCATGTGAATGCCCAGTGATAGCAACTGATGCA 7692

Qy        895 GGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATT 954
              || || |||   || ||  ||      | | ||    |  | ||  |   | ||   | |
Db       7693 GGGGGTGTTGCTGAGGTTCTTTCTAATGCTGATTGGCTGGTTCCAATTGCTGATAGCAGT 7752

Qy        955 CTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATA 1014
                || || |||||||| | ||||    ||    | ||   |   | |   |||   ||
Db       7753 AAATTAGCTGAGAAAATTGATGAGTTTTTTTTGCTTGACTCTTCAGAAGTTAAAGATATC 7812

Qy       1015 GGTATGAAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAG 1074
                |||||| | |||  ||  ||    | | ||||| |||    | ||| | | |
Db       7813 AAAGCGAAAAACAAAGACCATTGTGAGAACCAATTTTCCATTGGGGCTATAATTAATAAT 7872

Qy       1075 TGGGAGCGCTTATAT 1089
                |||  |   | | |||
Db       7873 TGGTATGCCATTTAT 7887


RESULT 12
CR626927_30/c
WPCOMMENT
  Sequence split into 52 fragments  LOCUS CR626927 Accession CR626927
     Fragment Name      Begin      End
     CR626927_00            1    110000
     CR626927_01       100001    210000
     CR626927_02       200001    310000
     CR626927_03       300001    410000
     CR626927_04       400001    510000
     CR626927_05       500001    610000
     CR626927_06       600001    710000
     CR626927_07       700001    810000
     CR626927_08       800001    910000
     CR626927_09       900001   1010000
     CR626927_10      1000001   1110000
     CR626927_11      1100001   1210000
     CR626927_12      1200001   1310000
     CR626927_13      1300001   1410000
     CR626927_14      1400001   1510000
     CR626927_15      1500001   1610000
     CR626927_16      1600001   1710000
     CR626927_17      1700001   1810000
     CR626927_18      1800001   1910000
     CR626927_19      1900001   2010000
     CR626927_20      2000001   2110000
     CR626927_21      2100001   2210000
     CR626927_22      2200001   2310000
     CR626927_23      2300001   2410000
     CR626927_24      2400001   2510000
     CR626927_25      2500001   2610000
     CR626927_26      2600001   2710000
     CR626927_27      2700001   2810000
     CR626927_28      2800001   2910000
```

CR626927_29	2900001	3010000
CR626927_30	3000001	3110000
CR626927_31	3100001	3210000
CR626927_32	3200001	3310000
CR626927_33	3300001	3410000
CR626927_34	3400001	3510000
CR626927_35	3500001	3610000
CR626927_36	3600001	3710000
CR626927_37	3700001	3810000
CR626927_38	3800001	3910000
CR626927_39	3900001	4010000
CR626927_40	4000001	4110000
CR626927_41	4100001	4210000
CR626927_42	4200001	4310000
CR626927_43	4300001	4410000
CR626927_44	4400001	4510000
CR626927_45	4500001	4610000
CR626927_46	4600001	4710000
CR626927_47	4700001	4810000
CR626927_48	4800001	4910000
CR626927_49	4900001	5010000
CR626927_50	5000001	5110000
CR626927_51	5100001	5205140

Continuation (31 of 52) of CR626927 from base 3000001 (CR626927 Bacteroides fragilis NCT

Query Match 19.1%; Score 214.4; DB 1; Length 110000;
Best Local Similarity 52.0%; Pred. No. 1.1e-25;
Matches 538; Conservative 0; Mismatches 481; Indels 15; Gaps 2;

```
Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
Db      20610 ATGAATATATTGTATATAATTACAGGTCTTGGTATTGGAGGTGCAGAAAAAGTAACGGTT 20551

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
Db      20550 GATCTTGCTAATTTAATGCAAGATAAGGGTAATAATGTTGCTATACTTTATCTATATGGT 20491

Qy        121 ATATCTGAGGTAAAGCC------AACACAAAATATTAATATTCATTATGTTAATATGGAT 174
Db      20490 GATAATCAATATACATCCCTTATATACAAATCTATATTTGTAAAGGGACTGTCTATGTCT 20431

Qy        175 AAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTA 234
Db      20430 AAAACACCAAAGAGCTTTTATTTGTCATTGATAAAAGCAAGGGAGATTTTGAAAAAGTTT 20371

Qy        235 AAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGG 294
Db      20370 CGGCCAGATGTTGTTCATGCACAAATGTTTCATGCTAATATTTTTGCTCGTTTTCTCCGT 20311

Qy        295 ATGCTGATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAAT 354
Db      20310 TTTTTTTATAAAATACCTCTTTTAATATGTACCGAACATAGTAAAAATATTGAGAGTAAG 20251

Qy        355 GCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGT 414
Db      20250 ATAAGAATGAGACTTTATCAGATAACAGATTTTTTGTCTGATGTAAATACCAATGTATCA 20191

Qy        415 AAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAG 474
Db      20190 AGAGAAGCAACAGAATATTTTATTTCTCATCATGCCTTTTCTGATTTAAAATCATTTCCA 20131

Qy        475 ATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACG 534
Db      20130 ATGTATAATGGAATATATTTGGAAAAATTTGAGAAAAGTAATGAGAGTAAACTTGTTTTG 20071

Qy        535 CGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTT 594
Db      20070 AAAAATAAAATTGGATTAGAAAATAATGAATTTGTTTTTTTAAATGTTGGAAGATTAGTA 20011

Qy        595 GAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACA 654
Db      20010 CAGGCTAAAGATCAAACTTCATTGTTGTTGGCTTTTTCTGATTTTTTAAAAACATATAAA 19951

Qy        655 TCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTA 714
Db      19950 TTGAATGCCAA---------ATTGGTAATTGTTGGAGAGGGACCATTGAGAGATAACTTA 19900

Qy        715 TTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGT 774
Db      19899 TGTAATTATATTAATAACTTGGATATTCGGGAAAATGTTATTCTTGCCGGATTTCATTCG 19840

Qy        775 GATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGT 834
Db      19839 AATGTTTCTGAATATTATAATGCGGCAGATTGCTTTGTACTATCCTCTCTTTGGGAAGGA 19780

Qy        835 TTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCT 894
Db      19779 TTTGGTATAGTTCTAGCAGAAGCTATGAGTTGTGGACTTCCTGTTATAACTACTGATGCT 19720

Qy        895 GGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATT 954
Db      19719 GGAGGCTGCAGGGAAGTAGTAGAAGATTCTCGCTTTGTAATTTCTTTGCAAAATCCTCAA 19660

Qy        955 CTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATA 1014
Db      19659 GAAATAACTTTAAAAATGAAAGAAATTTTTGATATGTCTTTGGAAGAAAGAGATATGTTG 19600

Qy       1015 GGTATGAAAAATAG 1028
Db      19599 GGAAATATGAATAG 19586
```

RESULT 13
AR318930
LOCUS AR318930 1077 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 1480 from patent US 6562958.
ACCESSION AR318930
VERSION AR318930.1 GI:33700033
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1077)
AUTHORS Breton,G. and Bush,D.
TITLE Nucleic acid and amino acid sequences relating to Acinetobacter baumannii for diagnostics and therapeutics
JOURNAL Patent: US 6562958-A 1480 13-MAY-2003;
Genome Therapeutics Corporation; Waltham, MA
FEATURES Location/Qualifiers
source 1. .1077
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Query Match 17.9%; Score 201.8; DB 6; Length 1077;
Best Local Similarity 52.8%; Pred. No. 4.8e-23;
Matches 493; Conservative 0; Mismatches 422; Indels 18; Gaps 2;

```
Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
Db         13 ATGAAAGTTATGTTGTTTATTACTGGTTTAGGCATGGGAGGAGCCGAGAAAGTTGTTTGC 72

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
Db         73 GAATTAGCTGATAAATTATATGAGTTGGGACATGAAGTCTGTATTGTATATTTTACGGG- 131

Qy        121 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 180
Db        132 -----TGAAATTATAAGACGACCGAAGAATAATATCAATATAATTTATATACCACTAAAA 186

Qy        181 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 240
Db        187 ATTAGTTCACTTATTTTTAGTGTATATTCAGTTATGAAACTAGTTGATCAATGGAAACCT 246

Qy        241 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 300
```

```
Db        247 GATGTAATCCATTCACATATGTTTCATGCAAATATAATTGCAAGATTGGCAAAAATAAGA 306

Qy        301 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 360

Db        307 AGAAAAAATATTCGTATAATTTGTAGTTCGCATTCTAATTATGAAGGTGGAGCATTACGT 366

Qy        361 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 420

Db        367 ATGATGATGTATGCTATGACTGAAAAATTATGTGATGTACATACAAATGTAAGTTCTAAT 426

Qy        421 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 480

Db        427 GCTGCTACTGCATTAATGCATGCTGGAGCAGTAAGAAAAAGAAAAATTAAAGCTGTTTAT 486

Qy        481 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540

Db        487 AATGGTATTAATATAAAAGACTATCATTTTGATGAAGTAGAAAACAAGTACTTACATAAA 546

Qy        541 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 600

Db        547 GCGTTTAACTTGAAAAATGAAGCTAAAACTATACTTGCAATAGGAAGGATAGATACTCCT 606

Qy        601 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 660

Db        607 AAAGATTATCCAAATTTATTAAAAGCTTTCAAAATAATTTATGAGATAGAC--------- 657

Qy        661 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 720

Db        658 ---AATAGTGTTCATCTATTTATAGCTGGAAATGGGCCAAAAAAAGAAGAAATCATGCTA 714

Qy        721 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 780

Db        715 TTGGCTAATAATTTAGGTATTAATAGCAATGTTAAATTTTTAGGCATACGGAATGATATT 774

Qy        781 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 840

Db        775 CCTAAGCTATTAAATTCATGTGATTTATTTATTTCATCGTCTGCATGGGAAGGATTTGGA 834

Qy        841 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 900

Db        835 CTAGCTATATTAGAGGCGATGATTTGTAAAAAACCTATTGTAGCAACTGAAACTGATGGT 894

Qy        901 GTTAAAGAAGTCGTTGGACCTCATAATGATGTT 933

Db        895 GCCAAAGAATTATTAAATAATGAAAATTTGGTT 927


RESULT 14
AP006841_17
WPCOMMENT
  Sequence split into 53 fragments  LOCUS AP006841 Accession AP006841
     Fragment Name      Begin     End
     AP006841_00            1    110000
     AP006841_01       100001    210000
     AP006841_02       200001    310000
     AP006841_03       300001    410000
     AP006841_04       400001    510000
     AP006841_05       500001    610000
     AP006841_06       600001    710000
     AP006841_07       700001    810000
     AP006841_08       800001    910000
     AP006841_09       900001   1010000
     AP006841_10      1000001   1110000
     AP006841_11      1100001   1210000
     AP006841_12      1200001   1310000
     AP006841_13      1300001   1410000
     AP006841_14      1400001   1510000
     AP006841_15      1500001   1610000
     AP006841_16      1600001   1710000
     AP006841_17      1700001   1810000
     AP006841_18      1800001   1910000
     AP006841_19      1900001   2010000
     AP006841_20      2000001   2110000
     AP006841_21      2100001   2210000
     AP006841_22      2200001   2310000
     AP006841_23      2300001   2410000
     AP006841_24      2400001   2510000
     AP006841_25      2500001   2610000
     AP006841_26      2600001   2710000
     AP006841_27      2700001   2810000
     AP006841_28      2800001   2910000
     AP006841_29      2900001   3010000
     AP006841_30      3000001   3110000
     AP006841_31      3100001   3210000
     AP006841_32      3200001   3310000
     AP006841_33      3300001   3410000
     AP006841_34      3400001   3510000
     AP006841_35      3500001   3610000
     AP006841_36      3600001   3710000
     AP006841_37      3700001   3810000
     AP006841_38      3800001   3910000
     AP006841_39      3900001   4010000
     AP006841_40      4000001   4110000
     AP006841_41      4100001   4210000
     AP006841_42      4200001   4310000
     AP006841_43      4300001   4410000
     AP006841_44      4400001   4510000
     AP006841_45      4500001   4610000
     AP006841_46      4600001   4710000
     AP006841_47      4700001   4810000
     AP006841_48      4800001   4910000
     AP006841_49      4900001   5010000
     AP006841_50      5000001   5110000
     AP006841_51      5100001   5210000
     AP006841_52      5200001   5277274
Continuation (18 of 53) of AP006841 from base 1700001 (AP006841 Bacteroides fragilis YCH

  Query Match           17.7%;  Score 199;  DB 1;  Length 110000;
  Best Local Similarity   51.5%;  Pred. No. 3.5e-23;
  Matches  569;  Conservative    0;  Mismatches  515;  Indels   21;  Gaps    4;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60

Db      97015 ATGAATATTATATATGTTATTACGGGACTTTCTCTTGGAGGTGCTGAGACTATAACCGTT 97074

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120

Db      97075 TGTCTGGCCAATGAAATGAGGCGGCGGGGACATCATGTAGCTATCTTGTATTTGACTGGC 97134

Qy        121 -ATATCTGAGGTAAAGCCAACA-----CAAAATATTAATATTCATTATGTTAATATGGAT 174

Db      97135 GAGAACTGGTTTAAAGACCGCATTGATAGAGATATTATAGTTCAAAGTTTAGGATTGTCA 97194

Qy        175 AAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTA 234

Db      97195 AAGAATCCTTTCAGTTTGCTAAAAGCACAATTGAAAGCTAAACAGTTTCTTAAAGTTTTT 97254

Qy        235 AAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGG 294

Db      97255 TCTCCTCAAGTAGTCCATGGAAATATGGTTCATGCTAATATTTTTGTAAGATTATTACGT 97314

Qy        295 ATGCTGATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAAT 354

Db      97315 CTTCATTGTCATATTCCTCTTTTGATATCTACTGAGCACAATAAAAATATAGAAGGACGG 97374

Qy        355 GCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGT 414

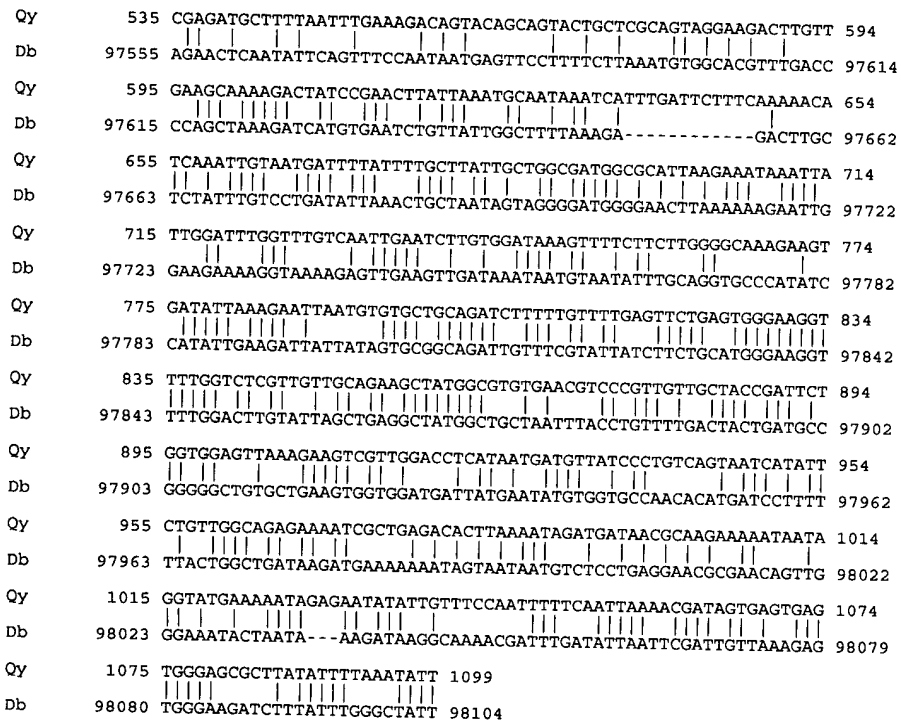
Db      97375 CTGAGGATGAAAATATATCAATATACAGATTTCTTGTCGGACATAAATACGAATGTTTCG 97434

Qy        415 AAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAG 474

Db      97435 AGGGAAGCTACAACTTATTTTATTCAGCAAAAAGCTTTCGGACGAAATAAAAGTATGACT 97494

Qy        475 ATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACG 534

Db      97495 GTTTATAACGGTATTGATCTTGTTCGCTTTGGAAGAAATGAAGATAAAAGATTGAAAATT 97554
```



RESULT 15
CR931708
LOCUS CR931708 19113 bp DNA linear BCT 30-JUN-2005
DEFINITION Streptococcus pneumoniae strain 1095/39 (serotype 36).
ACCESSION CR931708
VERSION CR931708.1 GI:68644250
KEYWORDS .
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 19113)
AUTHORS Bentley,S.D., Aanensen,D., Mavroidi,A., Saunders,D.,
Rabbinowitsch,E., Collins,M., Donaghue,K., Harris,D., Kaltoft,M.S.,
Murphy,L., Quail,M.A., Samual,G., Skovsted,I.C., Barrell,B.G.,
Reeves,P., Parkhill,J. and Spratt,B.G.
TITLE Genetic analysis of the capsular biosynthetic locus from all 90
serotypes of Streptococcus pneumoniae
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 19113)
AUTHORS Bentley,S.D.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-2004) Submitted on behalf of the Pathogen
Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, E-mail: sdb@sanger.ac.uk
COMMENT NOTE: This sequence was generated from a PCR product representing
the region from dexB to aliA and is not necessarily responsible for
the expressed capsule serotype. For a detailed description of how
CDS products were predicted see the associated publication.
FEATURES Location/Qualifiers
source 1. .19113
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/strain="1095/39"
/db_xref="taxon:1313"
gene <1. .137
/gene="dexB"
/locus_tag="SPC36_0001"
CDS <1. .137
/gene="dexB"
/locus_tag="SPC36_0001"
/codon_start=3
/transl_table=11
/protein_id="CAI34356.1"
/db_xref="GI:68644251"
/translation="SNEEQDLTVEGKVKSVLIENTLAQEVFEKQILVPWDAFCVEMTD
"
gene complement(284. .1659)
/gene="tnp"
/locus_tag="SPC36_0002"
/pseudo
CDS complement(join(284. .400,400. .1659))
/gene="tnp"
/locus_tag="SPC36_0002"
/note="member of homology group 0017 25"
/pseudo
/codon_start=1
/transl_table=11
/product="putative IS1202 transposase"
misc_feature complement(676. .1176)
/gene="tnp"
/locus_tag="SPC36_0002"
/note="HMMPfam hit to PF00665, Integrase, catalytic
region, score 5.4e-19"
/pseudo
gene 1835. .3280
/gene="wzg"
/locus_tag="SPC36_0003"
CDS 1835. .3280
/gene="wzg"
/locus_tag="SPC36_0003"
/note="member of homology group 0000 90"
/codon_start=1
/transl_table=11
/product="integral membrane regulatory protein Wzg"
/protein_id="CAI34358.1"
/db_xref="GI:68644252"
/translation="MSRRFKKSRSQKVKRSVNIVLLTIYLLLVCFLLFLIFKYNILAF
RYLNLVVTALVLLVALVGLLLIIYKKAEKFTIFLLVFSILVSSVSLFAVQQFVGLTNR
LNATSNYSEYSISVAVLADSDIKNVTQLTSVTAPTGTDNENIQKLLADIKSSQNTDLT
VDQSSSYLAAYKSLIAGETKAIVLNSVFENIIESEYPDYASKIKKIYTKGFTKKVEAP
KTSKNQSFNIYVSGIDTYGPISSVSRSDVNILMTVNRDTKKILLTTTPRDAYVPIADG
GNNQKDKLTHAGIYGVDSSIHTLENLYGVDINYYVRLNFTSFLKMIDLLGGVDVHNDQ
EFSALHGKFHFPVGNVHLDSEQALGFVRERYSLADGDRDRGRNQQKVIVAIIQKLTST
EVLKNYSSILQGLQDSLQTNMPIETMIDLVNTQLESGGNYKVNSQDLKGTGRMDLPSY
AMPDSNLYVMEVDDSSLAVVKAAIQDVMEGR"
misc_feature 1835. .1951
/gene="wzg"
/locus_tag="SPC36_0003"
/note="Signal peptide predicted for SPC1724 by SignalP 2.0
HMM (Signal peptide probability 0.999) with cleavage site
probability 0.540 between residues 39 and 40"
misc_feature join(1883. .1942,1970. .2038,2057. .2125)
/gene="wzg"
/locus_tag="SPC36_0003"
/note="3 probable transmembrane helices predicted for
SPC1724 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"
misc_feature 2042. .2386
/gene="wzg"
/locus_tag="SPC36_0003"
/note="HMMPfam hit to PF02916, DNA polymerase processivity
factor, score 2.7e-62"
misc_feature 2564. .3007
/gene="wzg"

```
                     /locus_tag="SPC36_0003"
                     /note="HMMPfam hit to PF03816, Cell envelope-related
                     transcriptional attenuator, score 3.2e-70"
     gene            3282. .4013
                     /gene="wzh"
                     /locus_tag="SPC36_0004"
     CDS             3282. .4013
                     /gene="wzh"
                     /locus_tag="SPC36_0004"
                     /EC_number="3.1.3.48"
                     /note="member of homology group 0001 90"
                     /codon_start=1
                     /transl_table=11
                     /product="protein-tyrosine phosphatase Wzh"
                     /protein_id="CAI34359.1"
                     /db_xref="GI:68644253"
                     /translation="MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRR
                     KGMFETPEEKIAENFLQVREIAKEVADDLVIAYGAEIYYTLDALEKLEKKEIPTLNDS
                     RYALIEFSMHTSYREIHTGLSNILMLGITPVIAHIERYDALENNEKRVCELIDMGCYT
                     QINSYHVLKPKFFGEKYKFMKKRARYFLERDLVHVVASDMHNLDSRPPHMAEAYDLVT
                     QKYGEAKAQELFVDNPRKIIMDQLI"
     misc_feature    3603. .3908
                     /gene="wzh"
                     /locus_tag="SPC36_0004"
                     /note="HMMPfam hit to PF02811, PHP, C-terminal, score
                     1.9e-10"
     gene            4022. .4714
                     /gene="wzd"
                     /locus_tag="SPC36_0005"
     CDS             4022. .4714
                     /gene="wzd"
                     /locus_tag="SPC36_0005"
                     /note="member of homology group 0002 90"
                     /codon_start=1
                     /transl_table=11
                     /product="capsular polysaccharide biosynthesis protein
                     Wzd"
                     /protein_id="CAI34360.1"
                     /db_xref="GI:68644254"
                     /translation="MKEQNTLEIDVLQLFRILWKRKLVILLVAIITSSVAFAYSTFVI
                     KPEFTSTTRIYVVNRNQEEKSGLTNQDLQAGSYLVKDYREIILSQDVLEEVISDLKLD
                     LTPKGLANKIKVTVPVDTRIVSVSVNDRVPEEASRIANSLREVAAQKIISITRVSDVT
                     TLEEARPAISPSLPNIKRNTLIGFLAGGIGTSVIVFLLELLDTHVKRPEDIEDTLQMT
                     LLGVVPNLSKLK"
     misc_feature    4040. .4456
                     /gene="wzd"
                     /locus_tag="SPC36_0005"
                     /note="HMMPfam hit to PF02706, Lipopolysaccharide
                     biosynthesis, score 1.2e-53"
     misc_feature    join(4088. .4153,4556. .4624)
                     /gene="wzd"
                     /locus_tag="SPC36_0005"
                     /note="2 probable transmembrane helices predicted for
                     SPC1726 by TMHMM2.0 at aa 23-44 and 179-201"
     misc_feature    4562. .4639
                     /gene="wzd"
                     /locus_tag="SPC36_0005"
                     /note="PS00217 Sugar transport proteins signature 2."
     gene            4724. .5407
                     /gene="wze"
                     /locus_tag="SPC36_0006"
     CDS             4724. .5407
                     /gene="wze"
                     /locus_tag="SPC36_0006"
                     /EC_number="2.7.1.112"
                     /note="member of homology group 0003 90"
                     /codon_start=1
                     /transl_table=11
                     /product="tyrosine-protein kinase Wze"
                     /protein_id="CAI34361.1"
                     /db_xref="GI:68644255"
                     /translation="MPTLEIAQKKLEFIKKAEEYYNALCTNIQLSGDKLKVISVTSVN
                     PGEGKTTTSVNIARSFARAGYKTLLIDGDTRNSVISGVFKSREKITGLTEFLSGTADL
                     SHGLCDTNIENLFVIQSGSVSPNPTALLQSKNFNDMIETLRKYFAYIIVDTAPIGIVI
                     DAAIITQKCDASILVTATGEVNKRDVQKAKQQLEQTGKLFLGVVLNKLDISVDKYGVY
                     GSYGNYGKK"
     gene            5423. .6790
                     /gene="wchA"
                     /locus_tag="SPC36_0007"
     CDS             5423. .6790
                     /gene="wchA"
                     /locus_tag="SPC36_0007"
                     /note="member of homology group 0005 66"
                     /codon_start=1
                     /transl_table=11
                     /product="undecaprenylphosphate glucosephosphotransferase
                     WchA (initial sugar transferase)"
                     /protein_id="CAI34362.1"
                     /db_xref="GI:68644256"
                     /translation="MDEKGLKIFLAVLQSIIVILLVYFLSFVRETELERSSMVILYLL
                     HFFVFYFSSYGNNFFKRGYLVEFNSTIRYIFFFAIAISVLNFFIAERFSISRRGMVYF
                     LTLEGISLYLLNFLVKKYWKHVFFNLKNSKKILLLTVTKNMEKVLDKLLESDELSWKL
                     VAVSVLDKSDFQHDKIPVIEKEKIIEFATHEVVDEVFVNLPGESYDIGEIISRFETMG
                     IDVTVNLKAFDKNLGRNKQIHEMEGLNVVTFSTNFYKTSHVISKRILDICGATIGLVL
                     FAIASLVLVPLIRKDGGPAIFAQTRIGKNGRHFTFYKFRSMRIDAEAIKEQLMDQNTM
                     QGGMFKMDNDPRITKIGRFIRKTSLDELPQFYNVLKGDMSLVGTRPPTVDEYEHYTPE
                     QKRRLSFKPGITGLWQVSGRSKIIDFDDVVKLDVAYIDNWTIWKDIEILLKTVKVVFM
                     RDGAK"

  Query Match             7.6%;  Score 85.4;  DB 1;  Length 19113;
  Best Local Similarity   51.4%;  Pred. No. 0.00014;
  Matches  197;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy        668 ATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTTT 727
              || ||  |||  |||| | ||| ||||| |   | |||| | || ||    ||   |||
Db       8934 ATATTCATTTATTTATCGTTGGTGATGGAGAGCTTAGAACTGAACTAGAAAATAAAGTTA 8993

Qy        728 GTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAAT 787
              |  | |||||  |  || |||||  | || || ||  | |||   ||||||   || |
Db       8994 AAGATCTCAATCTAGAAGAGAAAGTAATTTTTTTTGGCTATAGACAAGATATTGTTGAGT 9053

Qy        788 TAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCGTTG 847
                ||   |  |   ||| ||| ||| ||    || |  |  || || ||||| || ||||
Db       9054 GTATTAATAGTTTTGATTTTTGTGTCTTACCCTCAGTATTTGAGGGATTTGGCCTAGTTG 9113

Qy        848 TTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAGTTAAAG 907
              |   |||||  |   |  || |     |  |  |||| || |||  |   ||| | || |
Db       9114 CTATTGAAGCATTTATGAATGGAAAAACAATGATTGCAACAGATATTTCAGGATTAAATG 9173

Qy        908 AAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTGGCAGAGA 967
              |||| |||  |    | |||| |||  |   |     ||| ||  |        ||
Db       9174 AAGTTGTTACAAAAGAAAATGGTGTGTTGGTTCCGCTTAAAGATGCTAGAGCCTTAGCTT 9233

Qy        968 AAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATGAAAAATA 1027
                  ||||   ||    |  ||      |    |||| ||   ||| ||   |   ||
Db       9234 CTGCTATTGAGGGGCTGTCGACTGACTTGCAACTTAGAACAAAGTTAGCTACTCAGGCTA 9293

Qy       1028 GAGAATATATTGTTTCCAATTTT 1050
              | || ||  ||     || |||
Db       9294 AAAAAGATTATGAGGAGAAATTT 9316


Search completed: May 10, 2006, 20:34:27
Job time : 4089.52 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:42:47 ; Search time 525.166 Seconds
(without alignments)
14276.972 Million cell updates/sec

Title: US-09-423-093B-1_COPY_11821_12945
Perfect score: 1125
Sequence: 1 atgaagattgtttatataat.....gtaataatataattgattga 1125

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1125	100.0	14516	2	AAX06748	Aax06748 E. coli O
	2	1125	100.0	14516	3	AAZ56375	Aaz56375 Escherich
	3	245.8	21.8	1104	4	ABA89403	Aba89403 Escherich
	4	245.8	21.8	4604	3	AAA15182	Aaa15182 DNA encod
c	5	245.8	21.8	6927	4	ABA89401	Aba89401 Escherich
	6	229.4	20.4	13584	14	AEB64840	Aeb64840 Escherich
	7	201.8	17.9	1077	9	ADA30193	Ada30193 DNA encod
	8	81.8	7.3	17433	13	ADW38769	Adw38769 Shigella
	9	77.6	6.9	370	4	AAF79751	Aaf79751 E coli tr
	10	73.8	6.6	1143	8	ACA22859	Aca22859 Prokaryot
	11	73.8	6.6	1156	8	ACA21827	Aca21827 Prokaryot
	12	72.4	6.4	15796	11	ADN02599	Adn02599 Shigella
	13	65.6	5.8	1116	10	ABX05981	Abx05981 S. pneumo
	14	65.6	5.8	1128	8	ABZ42165	Abz42165 Streptoco
	15	65.6	5.8	12110	14	AEA33399	Aea33399 S. boydii
c	16	65.6	5.8	19250	2	AAV52168	Aav52168 Streptoco
	17	65.6	5.8	110000	10	ABS56454_02	Continuation (3 of
	18	65.4	5.8	14245	3	AAZ30356	Aaz30356 Nucleotid
	19	65.4	5.8	14245	3	AAZ45259	Aaz45259 DNA encod
c	20	63.6	5.7	3262	13	ADT05373	Adt05373 Haemophil
c	21	63.6	5.7	308766	13	ADT05738	Adt05738 Haemophil
	22	62.6	5.6	1092	8	ACA28175	Aca28175 Prokaryot
	23	62.4	5.5	110000	10	ADF77343_11	Continuation (12 o
	24	61.8	5.5	807	8	ACF75141	Acf75141 Staphyloc
	25	60.6	5.4	1101	13	ADS58469	Ads58469 Bacterial
	26	60	5.3	4985	6	ABQ75107	Abq75107 Anopheles
	27	60	5.3	4985	10	ACF79720	Acf79720 Mosquito
	28	58.6	5.2	12434	14	AEC15936	Aec15936 E. coli O
c	29	58	5.2	110000	2	AAT42063_17	Continuation (18 o
	30	57.2	5.1	1095	11	ACH95138	Ach95138 Klebsiell
c	31	56.8	5.0	6474	2	AAV52288	Aav52288 Streptoco
	32	56.8	5.0	110000	10	ABS56454_00	Abs56454 Streptoco
	33	56	5.0	17623	12	ADF77651	Adf77651 Francisel
	34	56	5.0	22401	12	ADF77691	Adf77691 Francisel
c	35	55.8	5.0	2000	11	ACL37108	Acl37108 Rice stre
c	36	55.4	4.9	351	9	ADB09599	Adb09599 Alloiococ
	37	55.4	4.9	1185	9	ADB09595	Adb09595 Alloiococ
	38	55.4	4.9	1185	9	ADB09597	Adb09597 Alloiococ
c	39	55.4	4.9	110000	9	ADB12064_09	Continuation (10 o
	40	55.2	4.9	525	13	ADK45674	Adk45674 Streptoco
	41	55.2	4.9	660	13	ADR92488	Adr92488 Novel S.
	42	55.2	4.9	660	14	AEA56358	Aea56358 Streptoco
	43	54.8	4.9	12220	14	AEA88828	Aea88828 Escherich
	44	54	4.8	453	10	ABX05746	Abx05746 S. pneumo
	45	54	4.8	2000	11	ACL37108	Acl37108 Rice stre

ALIGNMENTS

RESULT 1
AAX06748
ID AAX06748 standard; DNA; 14516 BP.
XX
AC AAX06748;
XX
DT 26-APR-1999 (first entry)
XX
DE E. coli O111 antigen gene cluster.
XX
KW O antigen; O111 antigen; diagnosis; wbdH gene; wzx gene; wzy gene;
KW wbdM gene; rfb gene; transferase; polymerase; flippase; diarrhoea;
KW haemorrhagic colitis; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 739. .1932
FT /*tag= a
FT /product= "galactosyl_transferase"
FT /note= "wbdH gene (ORF1), this region is specifically
FT claimed in Claim 6, and encodes AAW88300"
FT CDS 1983. .3101
FT /*tag= b
FT /note= "gmd gene (ORF2); encodes AAW88301"
FT CDS 3174. .3623
FT /*tag= c
FT /note= "wbdI gene (ORF3); encodes AAW88302"
FT CDS 3657. .5105
FT /*tag= d
FT /note= "manC gene (ORF4); encodes AAW88303"
FT CDS 5114. .6484
FT /*tag= e
FT /note= "manB gene (ORF5); encodes AAW88304"
FT CDS 6521. .7444
FT /*tag= f
FT /note= "wbdJ gene (ORF6); encodes AAW88305"
FT CDS 7441. .8607
FT /*tag= g
FT /note= "wbdK gene (ORF7); encodes AAW88306"
FT CDS 8646. .9911
FT /*tag= h

```
FT                   /product= "O_antigen_flippase"
FT                   /note= "wzx gene (ORF8), this region is specifically
FT                   claimed in Claim 6; encodes AAW88307"
FT   CDS             9901. .10953
FT                   /*tag=  i
FT                   /product= "O_antigen_polymerase"
FT                   /note= "wzy gene (ORF9), this region is specifically
FT                   claimed in Claim 6; encodes AAW88308"
FT   CDS             10931. .11824
FT                   /*tag=  j
FT                   /product= "colitose_or_glucose_transferase"
FT                   /note= "wbdL gene (ORF10); encodes AAW88309"
FT   CDS             11821. .12945
FT                   /*tag=  k
FT                   /product= "colitose or glucose transferase"
FT                   /note= "wbdM gene (ORF11), this region is specifically
FT                   claimed in Claim 6; encodes AAW88310"
FT   CDS             13124. .14516
FT                   /*tag=  l
FT                   /note= "gnd gene (partial sequence); encodes AAW88311"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;    98WO-AU000315.
XX
PR   01-MAY-1997;    97AU-00006545.
PR   22-JUL-1997;    97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88300, AAW88301, AAW88302, AAW88303, AAW88304, AAW88305,
DR   AAW88306, AAW88307, AAW88308, AAW88309, AAW88310, AAW88311.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 6; Fig 7; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Escherichia coli O111 O antigen. The gene cluster was
CC   isolated from a cosmid gene bank of E. coli M92 chromosomal DNA by
CC   screening with E. coli O111 antiserum. Open reading frames within the
CC   gene cluster were identified on the basis of homology to known sequences.
CC   The genes encode proteins (see AAW88300-11) involved the synthesis of
CC   sugars present in the polysaccharide antigen, and in the transport or
CC   processing of polysaccharide or oligosaccharide units. The use of nucleic
CC   acid molecules derived from particular assembly and transport genes,
CC   particularly wbd (transferase), wzx (flippase) and wzy (polymerase)
CC   genes, within O antigen gene clusters improves the specificity of methods
CC   for the detection and identification of O antigens, e.g. in testing food-
CC   or faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of enteropathogenic E. coli strains that cause
CC   diarrhoea and haemorrhagic colitis
XX
SQ   Sequence 14516 BP; 4582 A; 2221 C; 2971 G; 4742 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 1125;  DB 2;  Length 14516;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-228;
  Matches 1125;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11821 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 11880

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11881 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 11940

Qy        121 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11941 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 12000

Qy        181 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12001 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 12060

Qy        241 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12061 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 12120

Qy        301 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12121 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 12180

Qy        361 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12181 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 12240

Qy        421 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12241 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 12300

Qy        481 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12301 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 12360

Qy        541 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12361 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 12420

Qy        601 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12421 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 12480

Qy        661 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12481 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 12540

Qy        721 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12541 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 12600

Qy        781 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12601 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 12660

Qy        841 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12661 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 12720

Qy        901 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12721 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 12780

Qy        961 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12781 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 12840

Qy       1021 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12841 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 12900

Qy       1081 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 1125
              |||||||||||||||||||||||||||||||||||||||||||||
Db      12901 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 12945

RESULT 2
AAZ56375
ID   AAZ56375 standard; DNA; 14516 BP.
```

```
XX
AC   AAZ56375;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:45.
XX
KW   Flagellin; fliC; antigen; detection; ds.
XX
OS   Escherichia coli.
XX
PN   WO9961458-A1.
XX
PD   02-DEC-1999.
XX
PF   21-MAY-1999;   99WO-AU000385.
XX
PR   21-MAY-1998;   98AU-00003634.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 2000-072598/06.
XX
PT   Novel nucleic acid molecule useful for the detection of flagellated
PT   bacterial strains in food, feces, etc.
XX
PS   Claim 3; Page 216-221; 245pp; English.
XX
CC   AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all or
CC   part of an Escherichia coli flagellin protein except a protein expressed
CC   by E. coli H1, H7, H12 or H48 type strains. The present invention also
CC   describes a method of detecting the presence of E. coli of a particular H
CC   serotype in a sample, comprising specifically hybridising a nucleic acid,
CC   preferably at least a pair, derived from a flagellating gene, specific
CC   for a particular flagellin gene associated with the H serotype, to any
CC   E.coli in the sample which contain the gene, and detecting any hybridised
CC   molecules, identifying the presence of that serotype in the sample. (I)
CC   are useful for: (1) detecting the presence of E. coli of H serotype in a
CC   sample by hybridising at least one or a pair of (I) to any E. coli in the
CC   sample and detecting the hybridised nucleic acid molecules; and (2) for
CC   detecting the presence of both O and H-serotypes of E. coli by
CC   hybridising at least one or a pair of (I) to any E. coli present in the
CC   sample and detecting the hybridised nucleic acid molecules. (I) is
CC   particularly useful for detecting the combination of O and H antigen.
CC   Hybridised (I) when using at least one (I) is detected by southern blot
CC   analysis and, when using a pair of (I), is detected by polymerase chain
CC   reaction (PCR). AAZ56399 to AAZ56420 represent primers used in the
CC   exemplification of the present invention
XX
SQ   Sequence 14516 BP; 4582 A; 2221 C; 2971 G; 4742 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1125;  DB 3;  Length 14516;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-228;
  Matches 1125;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11821 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 11880

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11881 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 11940

Qy        121 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11941 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 12000

Qy        181 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12001 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 12060

Qy        241 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12061 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 12120

Qy        301 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12121 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 12180

Qy        361 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12181 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 12240

Qy        421 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12241 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 12300

Qy        481 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12301 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 12360

Qy        541 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12361 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 12420

Qy        601 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12421 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 12480

Qy        661 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12481 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 12540

Qy        721 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12541 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 12600

Qy        781 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12601 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 12660

Qy        841 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12661 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 12720

Qy        901 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12721 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 12780

Qy        961 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12781 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 12840

Qy       1021 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12841 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 12900

Qy       1081 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 1125
              |||||||||||||||||||||||||||||||||||||||||||||
Db      12901 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 12945


RESULT 3
ABA89403
ID   ABA89403 standard; DNA; 1104 BP.
XX
AC   ABA89403;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Escherichia coli polynucleotide SEQ ID NO 1321.
XX
KW   Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW   immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
```

KW systemic infection; non-diarrhoeal infection; septicaemia;
KW pyelonephritis; antibiotic resistance; ds.
XX
OS Escherichia coli.
XX
PN WO200166572-A2.
XX
PD 13-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-EP003445.
XX
PR 10-MAR-2000; 2000FR-00003145.
PR 02-FEB-2001; 2001FR-00001449.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX
DR WPI; 2001-550253/61.
XX
PT A library of DNA fragments of Escherichia coli strains for the phylogenic
PT determination of a given strain comprises polynucleotides of nature B2/D+
PT A-.
XX
PS Example 6; Fig 6; 646pp; English.
XX
CC The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and
CC encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC B2/D+A-. The polynucleotides have potential antiinflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicaemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more frequent
CC use of broad spectrum antibiotics
XX
SQ Sequence 1104 BP; 368 A; 146 C; 219 G; 371 T; 0 U; 0 Other;

Query Match 21.8%; Score 245.8; DB 4; Length 1104;
Best Local Similarity 55.8%; Pred. No. 2.4e-42;
Matches 550; Conservative 0; Mismatches 402; Indels 33; Gaps 3;

```
Qy    1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
Db    1 ATGAAAATACTATTTGTCATTACAGGTTTAGGCCTTGGAGGTGCTGAGAAGCAGGTTTGT 60

Qy   61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
Db   61 CTTTTAGCTGATAAATTAAGTTTAAGCGGGCACCATGTAAAGATTATTTCACTTGGACAT 120

Qy  121 ATATCTGAGGTAAAGCCAACAC---------AAAATATTAATATTCATTATGTTAATATG 171
Db  121 ATGTCTAATAATAAAGTCTTTCCTAGCGAAAATAATGTTAATGTCATTAATGTAAATATG 180

Qy  172 GATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCC 231
Db  181 TCAAAAAACATTTCTGGAGTTATAAAAGGTTGTGTCAGAATTAGAGATGTTATAGCTAAT 240

Qy  232 TTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATT 291
Db  241 TTCAAACCAGACATTGTACACAGTCATATGTTTCATGCAAACATTATCACTAGATTGTCT 300

Qy  292 AGGATGCTGATTCCAGCGGTGCC---CCTGATATGTACCGCACACAACAAAAATGAAGGT 348
Db  301 GTAATTGGAATCAAAAACAGACCTGGTATTATATCAACTGCACATAATAAAAATGAAGGT 360

Qy  349 GGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAAT 408
Db  361 GGGTATTTCAGAATGCTCACATATAGAATAACCGATTGTTTAAGTGATTGTTGTACAAAT 420

Qy  409 GTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATA 468
Db  421 GTTAGCAAAGAAGCAGTGGATGAGTTTTTACGGATAAAAGCCTTTAATCCCGCTAAAGCA 480

Qy  469 GTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAG 528
Db  481 ATTACTATGTATAATGGGATAGATACCAATAAATTTAAATTTGATTTATTGGCAAGGAGG 540

Qy  529 AAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGA 588
Db  541 GAAATTCGAGACGGTATTAATATAAAAAATGATGATATATTATTACTTGCTGCAGGTCGT 600

Qy  589 CTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCA 648
Db  601 TTAACGTTAGCTAAAGATTATCCTAATTTATTGAATGCAATGACTCTGCT---------- 650

Qy  649 AAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAAT 708
Db  651 ----------TCCTGAACACTTTAAACTTATTATTATTGGTGATGGTGAATTGCGTGAC 699

Qy  709 AAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAA 768
Db  700 GAAATTAATATGCTTATAAAAAAATTGCAATTATCTAATAGGGTGTCCTTGTTGGGAGTT 759

Qy  769 AGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGG 828
Db  760 AAAAAAAATATTGCTCCCTATTTTTCTGCATGTGATATTTTTGTTCTCTCTTCTCGTTGG 819

Qy  829 GAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACC 888
Db  820 GAAGGATTTGGATTAGTCGTGGCAGAAGCTATGTCATGTGAGCGAATTGTTGTTGGCACG 879

Qy  889 GATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAAT 948
Db  880 GATTCAGGGGGAGTAAGAGAAGTTATTGGTGACGATGATTTTCTTGTACCCATATCTGAT 939

Qy  949 CATATTCTGTTGGCAGAGAAAATCG 973
Db  940 TCAACACAACTTGCAAGCAAAATTG 964
```

RESULT 4
AAA15182
ID AAA15182 standard; DNA; 4604 BP.
XX
AC AAA15182;
XX
DT 04-SEP-2000 (first entry)
XX
DE DNA encoding Escherichia coli virulence proteins.
XX
KW Virulence protein; tatA; tatB; tatC; tatE; mdoG; creC; recG; yggN; eck1;
KW iroD; iroC; iroE; mtd2; ms1; vaccine; infection; Gram negative bacterium;
KW ss.
XX
OS Escherichia coli.
XX

FH	Key	Location/Qualifiers
FT	CDS	6. .746
FT		/*tag= a
FT		/product= "virulence protein"
FT		/note= "encodes AAY93229; no termination codon given"
FT	CDS	749. .1531
FT		/*tag= b
FT		/product= "virulence protein"
FT		/note= "encodes AAY93230; no termination codon given"
FT	CDS	1534. .2640
FT		/*tag= c
FT		/product= "virulence protein"
FT		/note= "encodes AAY93231"
FT	CDS	2982. .4085
FT		/*tag= d

```
FT                   /product= "virulence protein"
FT                   /note= "encodes AAY93232"
XX
PN   WO200028038-A2.
XX
PD   18-MAY-2000.
XX
PF   09-NOV-1999;   99WO-GB003721.
XX
PR   09-NOV-1998;   98GB-00024569.
PR   09-NOV-1998;   98GB-00024570.
PR   17-DEC-1998;   98GB-00027814.
PR   17-DEC-1998;   98GB-00027815.
PR   17-DEC-1998;   98GB-00027816.
PR   17-DEC-1998;   98GB-00027818.
PR   13-JAN-1999;   99GB-00000708.
PR   13-JAN-1999;   99GB-00000710.
PR   13-JAN-1999;   99GB-00000711.
PR   28-JAN-1999;   99GB-00001915.
XX
PA   (MICR-) MICROSCIENCE LTD.
XX
PI   Crooke HR,  Clarke EE,  Everest PH,  Dougan G,  Holden DW,  Shea JE;
PI   Feldman RG;
XX
DR   WPI; 2000-376550/32.
DR   P-PSDB; AAY93229, AAY93230, AAY93231, AAY93232.
XX
PT   Peptide encoded by an operon including genes from Escherichia coli for
PT   screening potential drugs, detecting virulence and treating conditions
PT   associated with infection by a Gram negative bacterium.
XX
PS   Example 9; Page 56-64; 122pp; English.
XX
CC   The present sequence encodes Escherichia coli virulence proteins. The
CC   specification describes virulence proteins which are encoded by an operon
CC   including tatA, tatB, tatC, tatE, mdoG, creC, recG, yggN, eck1, iroD,
CC   iroC, iroE, mtd2 or ms1-16 genes obtained from Escherichia coli K1. The
CC   virulence proteins and polynucleotides, and their vaccines are useful for
CC   screening potential drugs, for the detection of virulence, and for
CC   treating or preventing conditions associated with infection by a Gram
CC   negative bacterium particularly Escherichia coli
XX
SQ   Sequence 4604 BP; 1588 A; 640 C; 864 G; 1512 T; 0 U; 0 Other;

  Query Match             21.8%;  Score 245.8;  DB 3;  Length 4604;
  Best Local Similarity   55.8%;  Pred. No. 3.2e-42;
  Matches  550;  Conservative    0;  Mismatches  402;  Indels   33;  Gaps    3;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||| ||  | | | | || || ||  |       ||| || || ||  | |   |
Db       1534 ATGAAAATACTATTTGTCATTACAGGTTTAGGCCTTGGAGGTGCTGAGAAGCAGGTTTGT 1593

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
              |  |||||  ||  || |   | || |||||||  |||| || |||||||  ||      |
Db       1594 CTTTTAGCTGATAAATTAAGTTTAAGCGGGCACCATGTAAAGATTATTTCACTTGGACAT 1653

Qy        121 ATATCTGAGGTAAAGCCAACAC---------AAAATATTAATATTCATTATGTTAATATG 171
              || ||| |    ||         |          | ||| ||||| |   | ||||  ||||||
Db       1654 ATGTCTAATAATAAAGTCTTTCCTAGCGAAAATAATGTTAATGTCATTAATGTAAATATG 1713

Qy        172 GATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCC 231
              |||||  ||    |  || | | || ||   |   | | | | |  | || |
Db       1714 TCAAAAAACATTTCTGGAGTTATAAAAGGTTGTGTCAGAATTAGAGATGTTATAGCTAAT 1773

Qy        232 TTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATT 291
              || || ||||| ||  |||| ||||||||||||||||| || ||| | | | | ||   |
Db       1774 TTCAAACCAGACATTGTACACAGTCATATGTTTCATGCAAACATTATCACTAGATTGTCT 1833

Qy        292 AGGATGCTGATTCCAGCGGTGCC---CCTGATATGTACCGCACACAACAAAAATGAAGGT 348
              ||    ||   |      ||     | ||||  || ||||| || ||||||||||||
Db       1834 GTAATTGGAATCAAAAACAGACCTGGTATTATATCAACTGCACATAATAAAAATGAAGGT 1893

Qy        349 GGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAAT 408
              ||  ||   || ||| |    ||| || | |  |||| ||||  |  | |  |||||||
Db       1894 GGGTATTTCAGAATGCTCACATATAGAATAACCGATTGTTTAAGTGATTGTTGTACAAAT 1953

Qy        409 GTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATA 468
              || || ||||| || ||  | |||||| ||   | ||| ||       |    ||||  |
Db       1954 GTTAGCAAAGAAGCAGTGGATGAGTTTTTACGGATAAAAGCCTTTAATCCCGCTAAAGCA 2013

Qy        469 GTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAG 528
              |    ||    |||   ||  |||| ||||||||| | ||||||  |    |  || | |
Db       2014 ATTACTATGTATAATGGGATAGATACCAATAAATTTAAATTTGATTTATTGGCAAGGAGG 2073

Qy        529 AAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGA 588
              |||  ||||| | | ||||| | ||| |    |      | || | || || | |||  |
Db       2074 GAAATTCGAGACGGTATTAATATAAAAAATGATGATATATTATTACTTGCTGCAGGTCGT 2133

Qy        589 CTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCA 648
              |      ||| ||||| ||||| || ||||| |||||||| | ||   |
Db       2134 TTAACGTTAGCTAAAGATTATCCTAATTTATTGAATGCAATGACTCTGCT---------- 2183

Qy        649 AAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAAT 708
                      |  | |  | ||||    |  |||||  ||| ||||| | |||  |  |
Db       2184 ----------TCCTGAACACTTTAAACTTATTATTATTGGTGATGGTGAATTGCGTGAC 2232

Qy        709 AAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAA 768
              || |       |  |      ||||| |  |      |||  || | ||| ||||
Db       2233 GAAATTAATATGCTTATAAAAAAAATTGCAATTATCTAATAGGGTGTCCTTGTTGGGAGTT 2292

Qy        769 AGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGG 828
              | ||   |||||      |   | | |||    ||| |||||||| |   ||||   |||
Db       2293 AAAAAAAATATTGCTCCCTATTTTTCTGCATGTGATATTTTTGTTCTCTCTTCTCGTTGG 2352

Qy        829 GAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACC 888
              |||||  |||||  | || || |||||||||||| | |||||  ||    |||||||  ||
Db       2353 GAAGGATTTGGATTAGTCGTGGCAGAAGCTATGTCATGTGAGCGAATTGTTGTTGGCACG 2412

Qy        889 GATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAAT 948
              |||||  || |||||  | ||||||  ||||      || ||  | || | ||  |  | ||
Db       2413 GATTCAGGGGGAGTAAGAGAAGTTATTGGTGACGATGATTTTCTTGTACCCATATCTGAT 2472

Qy        949 CATATTCTGTTGGCAGAGAAAATCG 973
                |  |   | |||   |||||
Db       2473 TCAACACAACTTGCAAGCAAAATTG 2497


RESULT 5
ABA89401/c
ID   ABA89401 standard; DNA; 6927 BP.
XX
AC   ABA89401;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Escherichia coli polynucleotide SEQ ID NO 1318.
XX
KW   Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW   immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW   systemic infection; non-diarrhoeal infection; septicaemia;
KW   pyelonephritis; antibiotic resistance; ds.
XX
OS   Escherichia coli.
XX
PN   WO200166572-A2.
XX
PD   13-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-EP003445.
XX
PR   10-MAR-2000; 2000FR-00003145.
PR   02-FEB-2001; 2001FR-00001449.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
```

PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX
DR WPI; 2001-550253/61.
XX
PT A library of DNA fragments of Escherichia coli strains for the phylogenic
PT determination of a given strain comprises polynucleotides of nature B2/D+
PT A-.
XX
PS Example 6; Fig 6; 646pp; English.
XX
CC The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and
CC encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC B2/D+A-. The polynucleotides have potential antiinflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicaemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more frequent
CC use of broad spectrum antibiotics
XX
SQ Sequence 6927 BP; 2415 A; 1236 C; 950 G; 2326 T; 0 U; 0 Other;

Query Match 21.8%; Score 245.8; DB 4; Length 6927;
Best Local Similarity 55.8%; Pred. No. 3.5e-42;
Matches 550; Conservative 0; Mismatches 402; Indels 33; Gaps 3;

```
Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
Db       2665 ATGAAAATACTATTTGTCATTACAGGTTTAGGCCTTGGAGGTGCTGAGAAGCAGGTTTGT 2606

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
Db       2605 CTTTTAGCTGATAAATTAAGTTTAAGCGGGCACCATGTAAAGATTATTTCACTTGGACAT 2546

Qy        121 ATATCTGAGGTAAAGCCAACAC---------AAAATATTAATATTCATTATGTTAATATG 171
Db       2545 ATGTCTAATAATAAAGTCTTTCCTAGCGAAAATAATGTTAATGTCATTAATGTAAATATG 2486

Qy        172 GATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCC 231
Db       2485 TCAAAAAACATTTCTGGAGTTATAAAAGGTTGTGTCAGAATTAGAGATGTTATAGCTAAT 2426

Qy        232 TTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATT 291
Db       2425 TTCAAACCAGACATTGTACACAGTCATATGTTTCATGCAAACATTATCACTAGATTGTCT 2366

Qy        292 AGGATGCTGATTCCAGCGGTGCC---CCTGATATGTACCGCACACAACAAAAATGAAGGT 348
Db       2365 GTAATTGGAATCAAAAACAGACCTGGTATTATATCAACTGCACATAATAAAAATGAAGGT 2306

Qy        349 GGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAAT 408
Db       2305 GGGTATTTCAGAATGCTCACATATAGAATAACCGATTGTTTAAGTGATTGTTGTACAAAT 2246

Qy        409 GTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATA 468
Db       2245 GTTAGCAAAGAAGCAGTGGATGAGTTTTTACGGATAAAAGCCTTTAATCCCGCTAAAGCA 2186

Qy        469 GTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAG 528
Db       2185 ATTACTATGTATAATGGGATAGATACCAATAAATTTAAATTTGATTTATTGGCAAGGAGG 2126

Qy        529 AAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGA 588
Db       2125 GAAATTCGAGACGGTATTAATATAAAAAATGATGATATATTATTACTTGCTGCAGGTCGT 2066

Qy        589 CTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCA 648
Db       2065 TTAACGTTAGCTAAAGATTATCCTAATTTATTGAATGCAATGACTCTGCT---------- 2016
```

```
Qy         649 AAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAAT 708
Db        2015 ----------TCCTGAACACTTTAAACTTATTATTATTGGTGATGGTGAATTGCGTGAC 1967

Qy         709 AAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAA 768
Db        1966 GAAATTAATATGCTTATAAAAAAAATTGCAATTATCTAATAGGGTGTCCTTGTTGGGAGTT 1907

Qy         769 AGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGG 828
Db        1906 AAAAAAAATATTGCTCCCTATTTTTCTGCATGTGATATTTTTGTTCTCTCTTCTCGTTGG 1847

Qy         829 GAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACC 888
Db        1846 GAAGGATTTGGATTAGTCGTGGCAGAAGCTATGTCATGTGAGCGAATTGTTGTTGGCACG 1787

Qy         889 GATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAAT 948
Db        1786 GATTCAGGGGGAGTAAGAGAAGTTATTGGTGACGATGATTTTCTTGTACCCATATCTGAT 1727

Qy         949 CATATTCTGTTGGCAGAGAAAATCG 973
Db        1726 TCAACACAACTTGCAAGCAAAATTG 1702
```

RESULT 6
AEB64840
ID AEB64840 standard; DNA; 13584 BP.
XX
AC AEB64840;
XX
DT 22-SEP-2005 (first entry)
XX
DE Escherichia coli O142 O-antigen biosynthetic gene cluster.
XX
KW Microorganism detection; O-antigen biosynthetic gene cluster;
KW lipopolysaccharide antigen; gene; ds.
XX
OS Escherichia coli; O142.
XX

FH	Key	Location/Qualifiers
FT	CDS	1140. .2156
FT		/*tag= a
FT		/product= "Orf1 protein"
FT	CDS	2176. .3261
FT		/*tag= b
FT		/product= "RmlB protein (orf2)"
FT	CDS	3261. .4160
FT		/*tag= c
FT		/product= "RmlD protein (orf3)"
FT	CDS	4212. .5090
FT		/*tag= d
FT		/product= "RmlA protein (orf4)"
FT	CDS	5095. .5652
FT		/*tag= e
FT		/product= "RmlC protein (orf5)"
FT	CDS	5649. .6893
FT		/*tag= f
FT		/product= "Wzx protein (orf6)"
FT	CDS	6945. .8126
FT		/*tag= g
FT		/product= "Wzy protein (orf7)"
FT	CDS	8108. .9043
FT		/*tag= h
FT		/product= "Orf8 protein"
FT	CDS	9040. .9828
FT		/*tag= i
FT		/product= "Orf9 protein"
FT	CDS	9825. .10919
FT		/*tag= j
FT		/product= "Orf10 protein"
FT	CDS	11163. .12182

```
FT                   /*tag=  k
FT                   /product= "Gne protein"
XX
PN   CN1554761-A.
XX
PD   15-DEC-2004.
XX
PF   22-DEC-2003; 2003CN-01117865.
XX
PR   22-DEC-2003; 2003CN-01117865.
XX
PA   (UYNA-) UNIV NANKAI.
XX
PI   Wang L,  Kong Q,  Feng L;
XX
DR   WPI; 2005-215829/23.
XX
PT   O-antigen specific nucleotide of E. coli 0142 type.
XX
PS   Claim 1; SEQ ID NO 1; 30pp; Chinese.
XX
CC   The invention relates to the O-antigen biosynthetic gene cluster of
CC   Escherichia coli O142 (AEB64840) and to variants thereof. The O-antigen
CC   of E. coli O142 is a lipopolysaccharide antigen. The invention also
CC   relates to PCR primers (AEB64841-AEB64842) for the amplification of the O
CC   -antigen gene cluster, oligonucleotides derived from the
CC   glycosyltransferase and oligose unit processing genes of the O-antigen
CC   gene cluster, and a method for detecting E. coli O142 in the environment
CC   and in the human body using oligonucleotides of the invention. The
CC   present sequence represents the specifically claimed sequence of the E.
CC   coli O142 O-antigen biosynthetic gene cluster.
XX
SQ   Sequence 13584 BP; 4098 A; 2298 C; 2882 G; 4306 T; 0 U; 0 Other;

  Query Match             20.4%;  Score 229.4;  DB 14;  Length 13584;
  Best Local Similarity   53.1%;  Pred. No. 1.2e-38;
  Matches  575;  Conservative    0;  Mismatches  481;  Indels   27;  Gaps    3;

Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||| || || | | | || || ||  |      ||| || || ||| | |   |
Db       9825 ATGAATATAGTGTTTGTGATTACTGGATTGGGACTTGGGGGGGCTGAAAAGCAAGTATGC 9884

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
                || || || | | ||  |      | |||| |  |    ||  | |  || || ||
Db       9885 TTATTGGCTGATCGACTGGCTGAGGCAGAGCATCAAATATCGATAGTGTCCCTTACAGGA 9944

Qy        121 ATATCTGAGGTAAAGCCAACACAAAA---TATTAATATTCATTATGTTAATATGGATAAA 177
               || ||  || ||||||    | ||   |||| | ||| |  |  ||||||||||||||
Db       9945 GGATGTGTTGTTAAGCCAGATAATAAGCATATTGAAATTTACCACCTTAATATGGATAAA 10004

Qy        178 AATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAG 237
              |   | | ||| || |||  || | |  |  || ||   || || |||    |  | |||
Db      10005 AGCCTAATAAGTTTCTTTTTAGGTGTTATAAAATTACGTAATATTATTTCTACTATTAAG 10064

Qy        238 CCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATG 297
              || |||||  | ||||| |||||||| ||||| |||||  |   | | || | || | |
Db      10065 CCTGATATTGTGCATAGCCATATGTTCCATGCAAATATCATGTCTAGATTAAGTAAGTTA 10124

Qy        298 CTGATTCC---AGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAAT 354
              ||  | ||     |       || |||||||| ||||| | || | |||| || || |
Db      10125 CTATTGCCTTATTCCCATAAGCTCATATGTACAGCACATAGCAGATATGAGGGAGGGAGG 10184

Qy        355 GCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGT 414
                | ||| ||||||||||  | |  |||| ||| |  ||| | |||||| || |||
Db      10185 CTTCGTATGCTTTGTTATCGTTTAACAGATTATTTCAGTGATATAAATACAAACGTTAGT 10244

Qy        415 AAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAG 474
              ||||||||  |  | || | | |     | ||   |     |    | ||||    | |
Db      10245 AAAGAGGCGCTCGATGAATATGTTAATAATAAATATTTCTCAGCGACTAAATCTATTGTT 10304

Qy        475 ATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACG 534
               ||   |||    || ||||| | |||||| ||||   |||| ||     | |   | |
Db      10305 GTTTATAATGGCGTTGATACAGAAAAATTTAATTTCAGTATTGATAATCGTATTTATATA 10364

Qy        535 CGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTT 594
              ||||    ||   | | ||  ||  |       |  | |||   || || ||| |
Db      10365 AGAGAGCAATTGTCTATTAATAACCATGATAGGTTGATTCTCAGTGTTGGCAGATTAAAC 10424

Qy        595 GAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACA 654
                || ||||| ||||| ||  |  ||   ||| | |  | |||
Db      10425 CCTGCTAAAGATTATCCTAATCTGCTAGCAGCATTCATGCTTTT---------------- 10468

Qy        655 TCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTA 714
                    | |  | | ||   |  ||||   ||| || || |    |  |    |  |
Db      10469 -----ACCTGAGCACTATAAGCTTGTTATCATTGGTGAGGGAGACGTTCGTTCACAGATT 10523

Qy        715 TTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGT 774
                  |  | ||     |    || | | |       |||   ||  | ||      || |
Db      10524 GAACAGATAATTAAAGATCACAAGTTAGGGACGCGTGTTCAGTTACTTGGTAGTGTAAAT 10583

Qy        775 GATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGT 834
              || | || || |      | ||||   ||| | ||||| ||   ||| |  |||||||||
Db      10584 AATGTCAATGATTATTATTCTGCTTGTGATTTATTTGTCTTATCTTCGGCTTGGGAAGGT 10643

Qy        835 TTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCT 894
              ||| |  | ||||||   || || ||||| ||| ||||   || |||   ||| ||| |
Db      10644 TTTAGCTTAGTTGTTATCGAGGCCATGGCATGTCAACGCATCGCTGTATGTACTGATGCA 10703

Qy        895 GGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATT 954
              ||||||| |||||||   ||      || | || |  || | ||        |||
Db      10704 GGTGGAGTAAAAGAAGCTTTTACTGATCGTCATTACATTGTACCAACTTCAAATGCCGCA 10763

Qy        955 CTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATA 1014
              || ||    |||||   |||       ||   ||  |  | |    | |||   ||||
Db      10764 GCTTTAGCTCGCAAAATTATTGAAGTTGATATGTTAACTATTGATCGTAAAAATGAAATA 10823

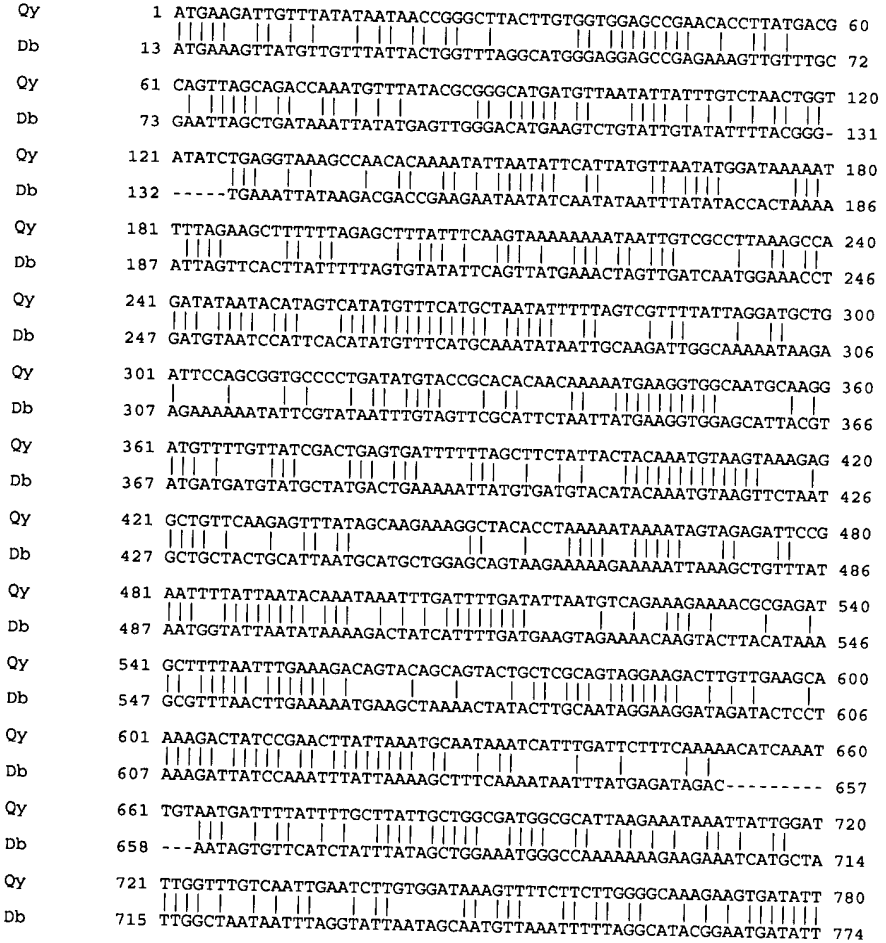
Qy       1015 GGTATGAAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAG 1074
                 |  | |||||  |  || |||||  ||| ||||||||||  | || || | | |
Db      10824 CAAAATGATAATAGGAATAATGTTGTTAACAAGTTTTCAATTAATGCTATTGTTAATCAT 10883

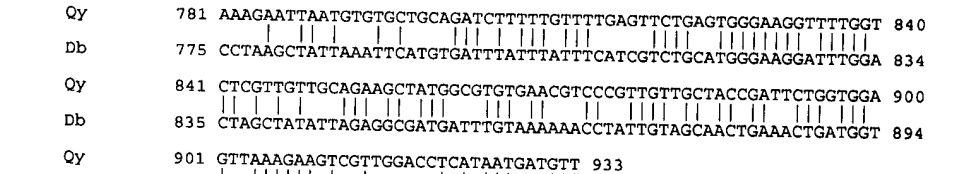
Qy       1075 TGG 1077
              |||
Db      10884 TGG 10886


RESULT 7
ADA30193
ID   ADA30193 standard; DNA; 1077 BP.
XX
AC   ADA30193;
XX
DT   20-NOV-2003  (first entry)
XX
DE   DNA encoding Acinetobacter baumannii protein #1480.
XX
KW   ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
KW   vaccine; plant biocontrol agent.
XX
OS   Acinetobacter baumannii.
XX
PN   US6562958-B1.
XX
PD   13-MAY-2003.
XX
PF   04-JUN-1999;   99US-00328352.
XX
PR   09-JUN-1998;   98US-0088701P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Breton G,  Bush D;
XX
DR   WPI; 2003-576092/54.
DR   P-PSDB; ADA34319.
XX
PT   New Acinetobacter baumanii proteins and nucleic acids, useful as reagents
```

PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
PS Example; SEQ ID NO 1480; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents DNA encoding an A. baumannii
CC protein.
XX
SQ Sequence 1077 BP; 411 A; 111 C; 201 G; 354 T; 0 U; 0 Other;

Query Match 17.9%; Score 201.8; DB 9; Length 1077;
Best Local Similarity 52.8%; Pred. No. 5e-33;
Matches 493; Conservative 0; Mismatches 422; Indels 18; Gaps 2;





RESULT 8
ADW38769
ID ADW38769 standard; DNA; 17433 BP.
XX
AC ADW38769;
XX
DT 24-MAR-2005 (first entry)
XX
DE Shigella dysenteriae O-antigen DNA SEQ ID NO:1.
XX
KW ds; antigen; O-antigen; gram-neg. bacterium.
XX
OS Shigella dysenteriae.
XX
PN CN1451663-A.
XX
PD 29-OCT-2003.
XX
PF 20-JAN-2003; 2003CN-00100358.
XX
PR 20-JAN-2003; 2003CN-00100358.
XX
PA (UYNA-) UNIV NANKAI.
XX
PI Wang L, Tao J, Feng L;
XX
DR WPI; 2004-110087/12.
XX
PT Nucleotide specific to O-antigen of shigella dysenteriae I and Bacillus
PT coli 0121.
XX
PS Claim 1; SEQ ID NO 1; 39pp; Chinese.
XX
CC The invention relates to a novel nucleotide specific to the O-antigen of
CC Shigella dysenteriae, represented in (ADW38769). The process for
CC obtaining the bacterial O-antigen gene cluster is also included. The O-
CC antigen can be used to detect Shigella dysenteriae.
XX
SQ Sequence 17433 BP; 5366 A; 3017 C; 3662 G; 5388 T; 0 U; 0 Other;

Query Match 7.3%; Score 81.8; DB 13; Length 17433;
Best Local Similarity 45.6%; Pred. No. 2.4e-07;
Matches 333; Conservative 0; Mismatches 392; Indels 6; Gaps 1;



```
Qy        628 ATAAATCATTTGATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCT 687
              |||          ||| || ||  ||  | |  |       ||  | ||  |   | |
Db      14384 GCTCTTCA------ACTTGCAGAAGAATTTAGTAATGTACGCTTCTTATTAGTCGGTCGT 14437

Qy        688 GGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGAT 747
              |||     | |    |  || | | || ||     | || ||  |      ||     |||
Db      14438 GGCTTGACCACCCAGAATAAAATAATAATGACACAGCTTAGTGTAAGCTGTCACTCAGAT 14497

Qy        748 AAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTT 807
              ||| || |  |  | ||  |  | || ||| ||   | ||   |   |||| | ||| ||
Db      14498 AAATTTATTCTTCTTGGTGAGCGTAGCGATGTTCCTGCATGTTTAAATGCTTCTGATATT 14557

Qy        808 TTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGT 867
              |||    |   ||| |     ||||| ||   |   ||  | | ||||||  |||||
Db      14558 TTTTGCCTTCATTCAGTAACAGAAGGCTTCCCTAATGTCTTAGGAGAAGCAATGGCAATG 14617

Qy        868 GAACGTCCCGTTGTTGCTACCGATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAAT 927
              | |  |||     ||  || | |   |||||  |     |     |  |||        | ||
Db      14618 GGAATTCCTTCCGTGACTTCAAACGTTGGTGATGCCGCATTTTTACTTGATAAGGTTGAT 14677

Qy        928 GATGTTATCCCTGTCAGTAATCATATTCTGTTGGCAGAGAAAATCGCTGAGACACTTAAA 987
               |||| | || | | | ||       | | || ||| ||||| |    ||      ||
Db      14678 TTTGTTGTTCCAGCCGGAAACTCGTATTTATTAGCACAGAAACTGCGGGAAGTTATTTTG 14737

Qy        988 ATAGATGATAACGCAAGAAAAATAATAGGTATGAAAAATAGAGAATATATTGTTTCCAAT 1047
               ||  |   ||     | |  |   |||||   | |   ||       ||||      |
Db      14738 TTAAGTCGAAAAAACCGCACTAAGTTAGGTCGTATATTAAGGCTGCGAATTGAGAATGAG 14797

Qy       1048 TTTTCAATTAAAACGATAGTGAGTGAGTGGGAGCGCTTATATTTTAAATATTCCAAGCGT 1107
              |||||  ||  | ||  |||  | | | |       |||||||   || ||    ||
Db      14798 TTTTCTATGGATACTGTAGCCAATAAATATTTATCCTTATATAAAAATTAAATGAAAAAG 14857

Qy       1108 AATAATATAAT 1118
              | |  | | ||
Db      14858 ATTGTTTTTAT 14868
```

RESULT 9
AAF79751
ID AAF79751 standard; DNA; 370 BP.
XX
AC AAF79751;
XX
DT 29-MAY-2001 (first entry)
XX
DE E coli trsE gene fragment.
XX
KW Virulence gene; K1 polysaccharide capsule; dgcA; dgcB; dgcC; dgcD; dgcE;
KW rnr; emrB; treB; pgi; adh; speA; frdA; rpoN; metJ; fnr; csgE; fimH; trsC;
KW trsE; antibacterial; infection; vaccine; attenuated bacterium; ds.
XX
OS Escherichia coli.
XX
PN WO200121655-A2.
XX
PD 29-MAR-2001.
XX
PF 22-SEP-2000; 2000WO-GB003647.
XX
PR 22-SEP-1999; 99EP-00307495.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Tang C;
XX
DR WPI; 2001-266066/27.
XX
PT Peptides useful as targets for antibacterial therapy, are encoded by
PT virulence genes from enteric bacteria that have a role in colonization
PT during infection.
XX
PS Claim 6; Page 17; 23pp; English.
XX
CC The present invention provides the coding sequences of several E. coli
CC proteins, including dgcA, dgcB, dgcC, dgcD, dgcE, rnr, emrB, treB, pgi,
CC adh, speA, frdA, rpoN, metJ, fnr, csgE, fimH, trsC and trsE. These can be
CC used in the diagnosis and treatment of bacterial infection, and disease
CC in prevention in the form of vaccines against E. coli. The present
CC sequence is one of the aforementioned coding sequences
XX
SQ Sequence 370 BP; 127 A; 46 C; 78 G; 119 T; 0 U; 0 Other;

Query Match 6.9%; Score 77.6; DB 4; Length 370;
Best Local Similarity 58.8%; Pred. No. 8.2e-07;
Matches 134; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```
Qy        397 ATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCT 456
              | |  ||||||||| || ||||| || ||  | ||||| ||   | ||| ||      |
Db         76 AGTTGTACAAATGTTAGCAAAGAAGCAGTGGATGAGTTTTTACGGATAAAAGCCTTTAAT 135

Qy        457 AAAAATAAAATAGTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATT 516
                  |||| | |    ||    |||   ||  |||| |||||||||| | |||||| |
Db        136 CCCGCTAAAGCAATTACTATGTATAATGGGATAGATACCAATAAATTTAAATTTGATTTA 195

Qy        517 AATGTCAGAAAGAAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTC 576
                | || | | |||  ||||| |   ||||| | ||| |   |     | || | ||
Db        196 TTGGCAAGGAGGGAAATTCGAGACGGGATTAATATAAAAAATGATGATATATTATTACTT 255

Qy        577 GCAGTAGGAAGACTTGTTGAAGCAAAAGACTATCCGAACTTATTAAAT 624
              || | |||  |  |      ||| ||||| ||||| || ||||| |||
Db        256 GCTGCAGGTCGTTTAACGTTAGCTAAAGATTATCCTAATTTATTGAAT 303
```

RESULT 10
ACA22859
ID ACA22859 standard; DNA; 1143 BP.
XX
AC ACA22859;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #4516.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Bacillus anthracis.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU18989.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 10729; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1143 BP; 381 A; 168 C; 252 G; 342 T; 0 U; 0 Other;

Query Match 6.6%; Score 73.8; DB 8; Length 1143;
Best Local Similarity 55.2%; Pred. No. 6.7e-06;
Matches 144; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

```
Qy        667 GATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTT 726
              |||   |  ||||||| ||| ||| |||||  ||  |      |   | ||  | || ||
Db        682 GATGCGAAGTTGCTTCTTGTTGGAGATGGACCAGAATTTTGCACGATTTTACAGTTAGTG 741

Qy        727 TGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAA 786
                | ||  || |||  |||   || || ||||| ||| |   |  | || ||  ||||
Db        742 AAAAATTTACATATTGAAGATCGCGTCTTATTCTTAGGGAAGCAAGATAATGTTGCAGAA 801

Qy        787 TTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCGTT 846
              || |      ||| |  |  |  |     || |||  |||  ||||||||||| |||
Db        802 CTACTTGCGATGAGTGATTTAATGTTGCTTTTATCAGAGAAGGAGAGTTTTGGTCTTGTT 861

Qy        847 GTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAGTTAAA 906
              |   |||||| ||||||||||      ||   | | |  ||       ||| ||  ||  |
Db        862 TTATTAGAAGCGATGGCGTGTGGTGTACCTTGTATCGGAACAAGGGTTGGAGGTATTCCA 921

Qy        907 GAAGTCGTTGGACCTCATAAT 927
              |||||| ||  || |  | ||
Db        922 GAAGTCATTCAACATGGTGAT 942
```

RESULT 11
ACA21827
ID ACA21827 standard; DNA; 1156 BP.
XX
AC ACA21827;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #3484.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Bacillus anthracis.
XX

PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU17957.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 9697; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1156 BP; 373 A; 181 C; 252 G; 350 T; 0 U; 0 Other;

Query Match 6.6%; Score 73.8; DB 8; Length 1156;
Best Local Similarity 55.2%; Pred. No. 6.7e-06;
Matches 144; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

```
Qy        667 GATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTT 726
              |||   |  ||||||| ||| ||| |||||  ||  |      |   | ||  | || ||
Db        703 GATGCGAAGTTGCTTCTTGTTGGAGATGGACCAGAATTTTGCACGATTTTACAGTTAGTG 762

Qy        727 TGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAA 786
                | ||  || |||  |||   || || ||||| ||| |   |  | || ||  ||||
Db        763 AAAAATTTACATATTGAAGATCGCGTCTTATTCTTAGGGAAGCAAGATAATGTTGCAGAA 822

Qy        787 TTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCGTT 846
```

```
Db        823 CTACTTGCGATGAGTGATTTAATGTTGCTTTTATCAGAGAAGGAGAGTTTTGGTCTTGTT 882

Qy        847 GTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAGTTAAA 906

Db        883 TTATTAGAAGCGATGGCGTGTGGTGTACCTTGTATCGGAACAAGGGTTGGAGGTATTCCA 942

Qy        907 GAAGTCGTTGGACCTCATAAT 927

Db        943 GAAGTCATTCAACATGGTGAT 963


RESULT 12
ADN02599
ID   ADN02599 standard; DNA; 15796 BP.
XX
AC   ADN02599;
XX
DT   17-JUN-2004  (first entry)
XX
DE   Shigella boydii serotype 7 O-antigen biosynthetic gene cluster (partial).
XX
KW   O-antigen biosynthetic gene cluster; O-antigen gene cluster; serotype 7;
KW   lipopolysaccharide antigen; detection; gene; ds.
XX
OS   Shigella boydii; serotype 7.
XX
FH   Key             Location/Qualifiers
FT   CDS             1160. .2007
FT                   /*tag=  a
FT                   /product= "FlmA protein (orf1)"
FT                   /note= "The start and stop positions featured are those
FT                   indicated by the inventors in the sequence shown on page
FT                   33-41. Decoding from the start codon given produces a 310
FT                   amino acid sequence with a stop codon ending at position
FT                   2092, rather than 2007 "
FT   CDS             2004. .3243
FT                   /*tag=  b
FT                   /product= "FlmB protein (orf2)"
FT                   /note= "The start and stop positions featured are those
FT                   indicated by the inventors in the sequence shown on page
FT                   33-41. Decoding from the start codon given produces a 9
FT                   amino acid sequence with a stop codon ending at position
FT                   2033 "
FT   CDS             3244. .3948
FT                   /*tag=  c
FT                   /product= "NeuA protein (orf3)"
FT   CDS             3932. .5032
FT                   /*tag=  d
FT                   /product= "FlmD protein (orf4)"
FT   CDS             5039. .5638
FT                   /*tag=  e
FT                   /product= "FlmH protein (orf5)"
FT   CDS             5662. .6717
FT                   /*tag=  f
FT                   /product= "NeuB protein (orf6)"
FT   CDS             6714. .7928
FT                   /*tag=  g
FT                   /product= "Wzx protein (orf7)"
FT   CDS             7936. .8877
FT                   /*tag=  h
FT                   /product= "Orf8 protein"
FT   CDS             8889. .10184
FT                   /*tag=  i
FT                   /product= "Wzy protein (orf9)"
FT                   /note= "The CDS start position given is that defined in
FT                   Claim 3, rather than that represented in the sequence
FT                   shown on pages 33-42 (8879). Decoding the protein from
FT                   position 8889 produces a 431 amino acid sequence with a
FT                   stop codon ending at position 10184, while decoding from
FT                   8879 produces a 5 amino acid sequence with a stop codon
FT                   ending at position 8896"
FT   CDS             10186. .11148
FT                   /*tag=  j
FT                   /product= "Orf10 protein"
FT   CDS             11355. .12272
FT                   /*tag=  k
FT                   /product= "Orf11 protein"
FT   CDS             12247. .13176
FT                   /*tag=  l
FT                   /product= "Orf12 protein"
FT   CDS             13193. .14344
FT                   /*tag=  m
FT                   /product= "Glf protein (orf13)"
XX
PN   CN1429832-A.
XX
PD   16-JUL-2003.
XX
PF   20-JAN-2003; 2003CN-00100536.
XX
PR   20-JAN-2003; 2003CN-00100536.
XX
PA   (UYNA-) UNIV NANKAI.
XX
PI   Wang L,  Tao J,  Feng L;
XX
DR   WPI; 2003-845847/79.
XX
PT   Nucleotide specific to O-antigen of shigella boydii 7.
XX
PS   Claim 1; SEQ ID NO 1; 41pp; Chinese.
XX
CC   The invention relates to the O-antigen biosynthetic gene cluster of
CC   Shigella boydii serotype 7 (ADN02599). The O-antigen of Shigella boydii
CC   is a lipopolysaccharide antigen. The invention also relates to PCR
CC   primers for the amplification of the 6-phosphogluconate dehydrogenase
CC   (gnd) or glycosyltransferase genes (ADN02591-ADN02596), a process for
CC   obtaining the O-antigen gene cluster, and a method for detecting Shigella
CC   boydii serotype 7 in the environment and in the human body. The present
CC   sequence represents the specifically claimed partial sequence of the
CC   Shigella boydii serotype 7 O-antigen biosynthetic gene cluster. Note: The
CC   features table gives CDS locations for 13 open reading frames (ORFs)
CC   which have mostly been obtained from the locations indicated in the
CC   sequence given on pages 33-42 of the specification, with the exception of
CC   the CDS start position for wzy (orf9) - see note attached to this
CC   feature. Also, the start and/or stop positions the inventors have
CC   indicated for flmA (orf1) and flmB (orf2) in this sequence may be
CC   incorrect - again see note attached to these features.
XX
SQ   Sequence 15796 BP; 4968 A; 2452 C; 3214 G; 5161 T; 0 U; 1 Other;

  Query Match             6.4%;  Score 72.4;  DB 11;  Length 15796;
  Best Local Similarity   49.3%;  Pred. No. 2.3e-05;
  Matches  218;  Conservative    0;  Mismatches  221;  Indels    3;  Gaps    1;

Qy        622 AATGCAATAAATCATTTGATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTT 681

Db      11772 ATTCTAATAAAATCCTTTGCTGAAATAGACTATTCTGTGAGGAAAGAATGGATCATGGAT 11831

Qy        682 ATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTT 741

Db      11832 ATTTATGGGGATGGAAATGAAAGAGCTGCCTTACAGAAATTAATTCATGATTTAAATATT 11891

Qy        742 GTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAATTAATGTGTGCTGCA 801

Db      11892 GAGGATTGTGTTAATTTATTGGGTAACAGTCGAGACCTTATGAACACTTATGCTGAATAT 11951

Qy        802 GATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATG 861

Db      11952 GATTTTTTTGTATTATCTTCAAGATTTGAAGGTTTCGGGATGGTTTTACTTGAAGCCATG 12011

Qy        862 GCGTGTGAACGTCCCGTTGTTGCTACCGATT---CTGGTGGAGTTAAAGAAGTCGTTGGA 918

Db      12012 AGTTGCGGTCTTCCTTGTATTGCCATCGATTGTCCTACAGGGCCTAGAGAAATACTTGAT 12071
```

```
Qy        919 CCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTGGCAGAGAAAATCGCTGAG 978
              ||  |||  || | |     || ||        |||  ||||  || ||
Db      12072 GGAGGCAAGTATGGGATACTTTGTGACAACCAATCAAATCTGGGGGAGAGTATTGCATCT 12131

Qy        979 ACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATGAAAAATAGAGAATATATT 1038
              | ||   |   | |||    ||  |  ||  || | | | | |   |  | |  |
Db      12132 CTAATTGTTAATAAGGATCTAAGAACTATGCTATCCGGCAAAAGCATTTTACGTAGTCAT 12191

Qy       1039 GTTTCCAATTTTTCAATTAAAA 1060
              | ||  ||| ||   |||| ||
Db      12192 GATTATAATATTCAGATTATAA 12213
```

RESULT 13
ABX05981
ID ABX05981 standard; DNA; 1116 BP.
XX
AC ABX05981;
XX
DT 27-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
DE S. pneumoniae type 4 strain coding region #269.
XX
KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KW ear infection; antiinflammatory; antibacterial; immunostimulant;
KW auditory; respiratory; gene therapy; vaccine.
XX
OS Streptococcus pneumoniae; type 4 strain.
XX
PN WO200277021-A2.
XX
PD 03-OCT-2002.
XX
PF 27-MAR-2002; 2002WO-IB002163.
XX
PR 27-MAR-2001; 2001GB-00007658.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Masignani V, Tettelin H, Fraser C;
XX
DR WPI; 2003-040579/03.
DR P-PSDB; ABU00702.
XX
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.
XX
PS Claim 6; SEQ ID NO 537; 56pp; English.
XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABS56454. Also included are an antibody which binds one of the proteins,
CC treating a patient by administering the protein, DNA or antibody (in a
CC composition), a kit comprising first and second primers, which are the
CC nucleic acid cited above or fragments between nucleotides 8-100 of a
CC sequence not defined in the specification, for amplifying a target
CC sequence contained within a Streptococcus nucleic acid sequence, where
CC the first primer is substantially complementary to the target sequence
CC and the second primer is substantially complementary to the complement of
CC the target sequence, and where the parts of the primers having
CC substantial complementarity define the termini of the target sequence to
CC be amplified, assay comprising contacting a test compound with the
CC protein, and determining whether the test compound binds to the protein
CC and a Streptococcus pneumoniae bacterium, where one or more genes
CC encoding the proteins has been rendered inactive. The proteins, nucleic
CC acid molecules, antibody and compositions are useful as medicaments for
CC treating or preventing a disease or infection due to streptococcus
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC media or ear infection. They are also useful in developing vaccines,
CC diagnostics and antibiotics. The methods are useful for identifying
CC immunodominant proteins. The present sequence is one of the 2489
CC identified coding region from the genomic sequence. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 1116 BP; 398 A; 135 C; 215 G; 368 T; 0 U; 0 Other;

  Query Match   5.8%; Score 65.6; DB 10; Length 1116;
  Best Local Similarity   47.8%; Pred. No. 0.00036;
  Matches 265; Conservative   0; Mismatches 274; Indels   15; Gaps   2;

```
Qy        416 AAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGA 475
              ||||  | | | |   |||     | |   || ||          |    ||||| | |   |
Db        437 AAGATTCGGCTAAGTGGTTATATGGAGAGAAAGCGTATAAAGACGGAAAAATTGAAATTA 496

Qy        476 TTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGC 535
              ||   |||   ||| ||   || || | | | |||        ||||  ||| | |||   |
Db        497 TTTTTAATGCAATTGATTTGAAAAAGTATCAATTTTTGTCAGATGTTAGAGAAAAATGTC 556

Qy        536 GAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTG 595
              |    |  ||  || ||  |   | ||      ||| |       |||   | ||    |
Db        557 GTAGAGAATTAGATGTGTCA---AATAAGTTCGTATTAGGAAATATAGCTCGCCTATCAG 613

Qy        596 AAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACAT 655
              |   |||| || |     | ||||| || |   ||||  |  | ||| |    ||| ||
Db        614 ATCAAAAAAACCAAAGTTATTTATTTAACGTTTTAAAAGAACTCATTTTAATCAAACCA- 672

Qy        656 CAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTAT 715
                      || ||||||| ||  | | |||  |||| |          | ||||||
Db        673 -----------AATGTTATTTTACTCCTAGTTGGTAATGGTGAGGATGAGCAGAAATTAA 721

Qy        716 TGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTG 775
              |    | ||   || | |||||          | ||  | || || |||  ||| | ||
Db        722 AACAGAAAGCTTTAGAACTAAATCTGACCCCATATGTGCTATTTTTAGGGAGAAGGACTG 781

Qy        776 ATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTT 835
              |||||   || ||| | | |||    ||| |||||  |   |     ||| | |  || || |
Db        782 ATATTTCTGATTTATTATCTGCGATGGATGTTTTTTTGCTTCCGTCTAAATATGAGGGGT 841

Qy        836 TTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTG 895
              |  | |   | | | ||| |||  ||| |  | |  |    || |  |  | ||| | |
Db        842 TGCCTGTTTCTCTAGTAGAGGCTCAGGCATCGGGATTACAAATTTTATCGTCAGATACAG 901

Qy        896 GTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTC 955
                  ||| ||  ||      | | |  |    | ||| |||    |  |  |
Db        902 TGACGCAAGAAGTAGATGTGACCAAAAACATTAGTTACTTACCTATCAACGAAGAGTCTG 961

Qy        956 TGTTGGCAGAGAAA 969
              |||||  |  ||||
Db        962 TGTTGCTATGGAAA 975
```

RESULT 14
ABZ42165
ID ABZ42165 standard; DNA; 1128 BP.
XX
AC ABZ42165;
XX
DT 04-MAR-2003 (first entry)
XX
DE Streptococcus pneumoniae polynucleotide SEQ ID NO 19.
XX
KW Streptococcus pneumoniae; infection; otitis media; antibacterial;
KW diagnosis; gene therapy; gene; ds.
XX
OS Streptococcus pneumoniae.
XX

PN WO200283855-A2.
XX
PD 24-OCT-2002.
XX
PF 12-APR-2002; 2002WO-US011524.
XX
PR 16-APR-2001; 2001US-0283948P.
PR 18-APR-2001; 2001US-0284443P.
XX
PA (AMCY) AMERICAN CYANAMID CO.
XX
PI Zagursky RJ, Masi AW, Green BA, Chakravarti DN, Russell DP;
PI Wooters JL;
XX
DR WPI; 2003-093010/08.
DR P-PSDB; ABP81317.
XX
PT New Streptococcus pneumoniae polynucleotides, useful for treating or
PT preventing S. pneumoniae infections, or non-systemic diseases, e.g.
PT otitis media, which are induced or exacerbated by S. pneumoniae.
XX
PS Claim 1; Page 193; 1091pp; English.
XX
CC The invention relates to isolated polynucleotides (ABZ72147-ABZ42522) of
CC a Streptococcus pneumoniae genomic sequence, a fragment or degenerate
CC variant of the polynucleotide or a nucleic acid sequence 95% identical to
CC one of the polynucleotides. The S. pneumoniae polynucleotides and encoded
CC polypeptides (ABP81299-ABP81674) are useful for treating or preventing S.
CC pneumoniae infections or non-systemic diseases, e.g. otitis media, which
CC are induced or exacerbated by S. pneumoniae. These are also useful for
CC detecting S. pneumoniae in a biological sample or diagnosing S.
CC pneumoniae infection in a subject. The polynucleotides have antibacterial
CC activity and are useful in gene therapy
XX
SQ Sequence 1128 BP; 404 A; 135 C; 220 G; 369 T; 0 U; 0 Other;

Query Match 5.8%; Score 65.6; DB 8; Length 1128;
Best Local Similarity 47.8%; Pred. No. 0.00036;
Matches 265; Conservative 0; Mismatches 274; Indels 15; Gaps 2;

```
Qy        416 AAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGA 475
              ||||  | | | |   |||    | |   || ||         |    ||||| | |   |
Db        446 AAGATTCGGCTAAGTGGTTATATGGAGAGAAAGCGTATAAAGACGGAAAAATTGAAATTA 505

Qy        476 TTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGC 535
              ||   |||   ||| ||   || || | | | |||       |||| ||| | |||   |
Db        506 TTTTTAATGCAATTGATTTGAAAAAGTATCAATTTTTGTCAGATGTTAGAGAAAAATGTC 565

Qy        536 GAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTG 595
              |    |  ||  || ||  |   | ||     ||| |         |||   | ||    |
Db        566 GTAGAGAATTAGATGTGTCA---AATAAGTTCGTATTAGGAAATATAGCTCGCCTATCAG 622

Qy        596 AAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACAT 655
              |   |||| || |     | ||||| || |   ||||  |  | ||| |    ||| ||
Db        623 ATCAAAAAAACCAAAGTTATTTATTTAACGTTTTAAAAGAACTCATTTTAATCAAACCA- 681

Qy        656 CAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTAT 715
                      || ||||||| ||  | | |||  |||| |          | ||||||
Db        682 ----------AATGTTATTTTACTCCTAGTTGGTAATGGTGAGGATGAGCAGAAATTAA 730

Qy        716 TGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTG 775
                |  | ||   || | |||||       | ||  | || || |||  ||| | ||
Db        731 AACAGAAAGCTTTAGAACTAAATCTGACCCCATATGTGCTATTTTTAGGGAGAAGGACTG 790

Qy        776 ATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTT 835
              |||||   || ||| | | |||    ||| ||||| |  |    ||| | |  || || |
Db        791 ATATTTCTGATTTATTATCTGCGATGGATGTTTTTTTGCTTCCGTCTAAATATGAGGGGT 850

Qy        836 TTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTG 895
              |   | |   | | | ||| |||  ||| |  | |    |   || |  |  | ||| | |
Db        851 TGCCTGTTTCTCTAGTAGAGGCTCAGGCATCGGGATTACAAATTTTATCGTCAGATACAG 910

Qy        896 GTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTC 955
```

```
                    ||| ||  ||        | | |  |    | ||| |||   |  |   |
Db        911 TGACGCAAGAAGTAGATGTGACCAAAAACATTAGTTACTTACCTATCAACGAAGAGTCTG 970

Qy        956 TGTTGGCAGAGAAA 969
              |||||  |  ||||
Db        971 TGTTGCTATGGAAA 984
```

RESULT 15
AEA33399
ID AEA33399 standard; DNA; 12110 BP.
XX
AC AEA33399;
XX
DT 28-JUL-2005 (first entry)
XX
DE S. boydii O-antigen gene, seq id 1.
XX
KW Microorganism detection; O-antigen; gene; ds.
XX
OS Shigella boydii; 18.
XX
PN CN1563066-A.
XX
PD 12-JAN-2005.
XX
PF 19-APR-2004; 2004CN-00019051.
XX
PR 19-APR-2004; 2004CN-00019051.
XX
PA (TIAN-) TIANJIN BIOCHIP TECH CO LTD.
XX
PI Wang L, Wang W, Feng L;
XX
DR WPI; 2005-307141/32.
XX
PT Nucleotide peculiar to O-antigen of Shigella boydii 18 useful for
PT controlling synthesis of O-antigen and detecting and identifying Shigella
PT in human body.
XX
PS Claim 1; SEQ ID NO 1; 33pp; Chinese.
XX
CC The invention provides a nucleotide which is specific to O-antigen of
CC Shigella boydii 18. It is the sequence of the gene cluster for
CC controlling the synthesis of the O-antigen. The sequence also includes
CC the oligonucleotide of the glycosyltransferase gene and oligosaccharide
CC unit treatment gene. The invention utilizes PCR to verify that the
CC oligonucleotide has high specificity to the O-antigen of Shigella boydii
CC 18, and also discloses the method for detecting and identifying Shigella
CC boydii 18 in the human body. The current sequence represents the S.
CC boydii O-antigen gene.
XX
SQ Sequence 12110 BP; 3823 A; 1968 C; 2484 G; 3835 T; 0 U; 0 Other;

Query Match 5.8%; Score 65.6; DB 14; Length 12110;
Best Local Similarity 47.7%; Pred. No. 0.0006;
Matches 225; Conservative 0; Mismatches 244; Indels 3; Gaps 1;

```
Qy        621 AAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCT 680
              | ||   | ||  | |||  |    | ||||  |   ||   | ||||||    | |
Db      11007 AGATCATAAAACACTTTTTTTAGCGTTAAAAGAACTTAAAGATTATGATTGGTCTCTTAG 11066

Qy        681 TATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCT 740
                | |  ||  | ||  |||||| | |  | |       | |  |    | || |   | |
Db      11067 CTTGGTCGGTAAAGGACCATTAATAGACCACTATAAAAAATATGCAATTGAACTTGGTAT 11126

Qy        741 TGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAATTAATGTGTGCTGC 800
              | ||      ||   ||  | ||  | ||||| ||| ||   || ||| | | ||
Db      11127 CTTCGAACGTATTAAATTTCTTGGAGAGAGAAGCGATGTTGCGGATTTATTATCTGAGAG 11186

Qy        801 AGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCGTTGTTGCAGAAGCTAT 860
              || | ||| |  | | |||| | |||||||||||  ||    | ||   ||||| ||
Db      11187 CAATGTATTTTTGCTTATTTCTAACTGGGAAGGTTTTCCTCGTAGTATTCTTGAAGCAAT 11246
```

```
Qy        861 GGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGG---AGTTAAAGAAGTCGTTGG 917
              |       |      || ||  ||||  | |||   || ||    |    ||  |||  ||
Db      11247 GAGAGCAGGCTTACCAGTAATTGCATCTGATGTAGGCGGGACATCGGAATCAGTTCTTAA 11306

Qy        918 ACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTGGCAGAGAAAATCGCTGA 977
                | | |   |  || |   |    | ||| || | |       | |  |||    |||
Db      11307 TAATAAAACAGGGTTTCTTATTTCTCGAAATGATGTCCATACTTTAAAAGAAAAGATTGA 11366

Qy        978 GACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATGAAAAATAGAGAATATAT 1037
              |   | |  | | ||||        | ||   ||| ||   | |   |||||||  |
Db      11367 AATCTTGATTAAAAATGAAGCACTTCGTAATGAAATGGGGCGGTATGCTAGAGAAATTTA 11426

Qy       1038 TGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAGCGCTTATAT 1089
              | |  | || ||| |  || |  | ||    | | ||   |     ||||||
Db      11427 TCTCGCAAACTTTACTTTTGATTCTATGTATAATAAGACAGTAGCATTATAT 11478
```

Search completed: May 10, 2006, 10:59:45
Job time : 529.166 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 11:00:02 ; Search time 3603.15 Seconds
(without alignments)
14608.170 Million cell updates/sec

Title: US-09-423-093B-1_COPY_11821_12945
Perfect score: 1125
Sequence: 1 atgaagattgtttatataat.....gtaataatataattgattga 1125

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

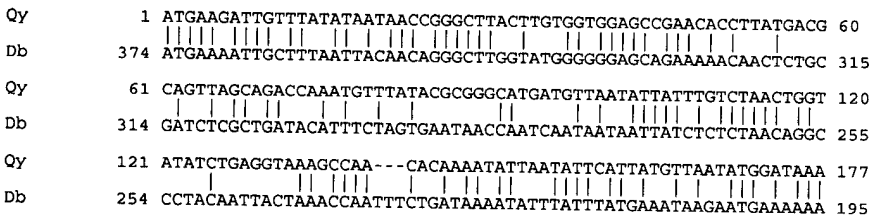
Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	90.8	8.1	671	9	AQ989576	AQ989576 Rfc00137
c 2	70.4	6.3	1101	10	CNS00EVL	AL069706 Drosophil
c 3	69.2	6.2	1101	10	CNS0039G	AL063921 Drosophil
c 4	66.6	5.9	781	10	BX145762	BX145762 Danio rer
5	66	5.9	1092	10	CNS020K7	AL175696 Tetraodon
c 6	63.6	5.7	1101	10	CNS00EPO	AL069493 Drosophil
7	63.4	5.6	1101	10	CNS0182P	AL108811 Drosophil
8	61.2	5.4	891	10	CW952279	CW952279 TcB38.1_B
9	61	5.4	755	10	BX154986	BX154986 Danio rer
c 10	60.6	5.4	1055	10	CL501343	CL501343 SAIL_696_
11	59.6	5.3	1101	10	CNS00EVL	AL069706 Drosophil
c 12	58.6	5.2	1563	9	CC206942	CC206942 CH261-191
c 13	58.4	5.2	987	10	CNS014PQ	AL104456 Drosophil
14	58.4	5.2	1101	10	CNS0039G	AL063921 Drosophil
15	58.2	5.2	1101	10	CNS0100X	AL098379 Drosophil
c 16	58.2	5.2	1211	10	AG349657	AG349657 Mus muscu
c 17	58	5.2	1037	1	AJ925726	AJ925726 AJ925726
c 18	57.8	5.1	587	9	BH873429	BH873429 hp45a07.b
c 19	57.8	5.1	709	10	CL531210	CL531210 trib004xb
c 20	57.6	5.1	735	10	CW960936	CW960936 AIAA-aac1
c 21	57.6	5.1	928	10	CNS00DKY	AL071865 Drosophil
c 22	57.4	5.1	1101	10	CNS0180W	AL108746 Drosophil
23	57	5.1	1101	10	CNS00LT2	AL078714 Drosophil
c 24	56.8	5.0	889	10	CW973509	CW973509 AIAA-aaa5
c 25	56.4	5.0	621	9	AQ941683	AQ941683 Sheared D
26	56.2	5.0	430	10	CG803196	CG803196 1118039H0
27	56.2	5.0	673	11	CNS06UWL	AL416395 T7 end of
28	56.2	5.0	1367	8	DN683496	DN683496 CGX30-C12
29	56	5.0	873	10	CNS024M5	AL180950 Tetraodon
30	56	5.0	996	10	CNS00FUH	AL071063 Drosophil
31	55.8	5.0	928	10	CNS00DKY	AL071865 Drosophil
c 32	55.8	5.0	1101	10	CNS003BB	AL064089 Drosophil
33	55.6	4.9	641	9	AQ946120	AQ946120 Sheared D
c 34	55.6	4.9	829	10	BX173672	BX173672 Danio rer
c 35	55.2	4.9	400	10	CG807323	CG807323 1118080G0
c 36	55	4.9	707	10	CG403581	CG403581 ZMMBBb024
37	55	4.9	889	10	CW973509	CW973509 AIAA-aaa5
c 38	55	4.9	902	10	CL475897	CL475897 SAIL_246_
39	55	4.9	1101	10	CNS0106X	AL098595 Drosophil
40	54.8	4.9	505	10	CW270542	CW270542 104_742_1
c 41	54.6	4.9	1092	10	CNS020K7	AL175696 Tetraodon
42	54.6	4.9	1101	10	CNS000D1	AL065414 Drosophil
c 43	54.4	4.8	471	10	CG807194	CG807194 1118078H0
c 44	54.4	4.8	700	9	AQ940248	AQ940248 Sheared D
c 45	54.4	4.8	1245	10	CG747942	CG747942 P041-3-G0

ALIGNMENTS

RESULT 1
AQ989576/c
LOCUS AQ989576 671 bp DNA linear GSS 14-AUG-2000
DEFINITION Rfc00137 Photorhabdus luminescens strain W14 M13 library Photorhabdus luminescens genomic clone PLG00137, genomic survey sequence.
ACCESSION AQ989576
VERSION AQ989576.1 GI:9648170
KEYWORDS GSS.
SOURCE Photorhabdus luminescens
ORGANISM Photorhabdus luminescens
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Photorhabdus.
REFERENCE 1 (bases 1 to 671)
AUTHORS ffrench-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T., Daborn,P.J., Bowen,D. and Blattner,F.R.
TITLE A genomic sample sequence of the entomopathogenic bacterium Photorhabdus luminescens W14: potential implications for virulence
JOURNAL Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
PUBMED 10919786
COMMENT Contact: ffrench-Constant RH
Department of Biology and Biochemistry
University of Bath
South Building, Bath BA2 7AY, UK
Tel: (44) 1225 826621
Fax: (44) 1225 826779
Email: bssrfc@bath.ac.uk
This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic Acids Res.
Seq primer: M13 Forward
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .671
/organism="Photorhabdus luminescens"
/mol_type="genomic DNA"
/strain="W14"
/db_xref="taxon:29488"
/clone="PLG00137"
/dev_stage="primary phase variant"
/clone_lib="Photorhabdus luminescens strain W14 M13 library"
/note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."

ORIGIN

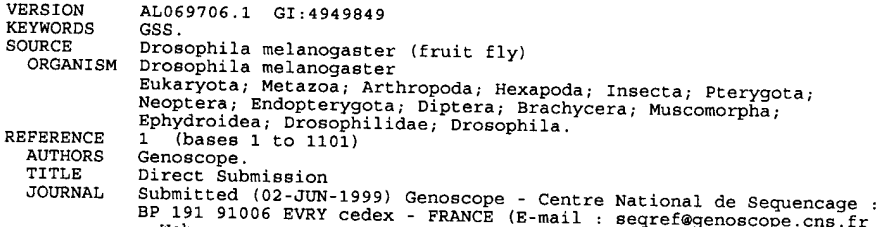
Query Match 8.1%; Score 90.8; DB 9; Length 671;
Best Local Similarity 54.5%; Pred. No. 2.3e-08;
Matches 204; Conservative 0; Mismatches 167; Indels 3; Gaps 1;



RESULT 2
CNS00EVL/c
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"

ORIGIN

Query Match 6.3%; Score 70.4; DB 10; Length 1101;
Best Local Similarity 33.2%; Pred. No. 0.00038;
Matches 207; Conservative 104; Mismatches 311; Indels 2; Gaps 1;



RESULT 3
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission

```
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             6.2%;  Score 69.2;  DB 10;  Length 1101;
  Best Local Similarity   16.7%;  Pred. No. 0.00067;
  Matches  100;  Conservative  259;  Mismatches  241;  Indels    0;  Gaps    0;

Qy        191 TTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCAGATATAATAC 250
              :| ::: :  ::: ::: :|:::::| :|: :|:::  |  :|| :   : ::||::|:
Db       1088 DTRKDDWDWTKWWTWWKDRADDRRWAGDADRWADDDGAGTWWTATWWWWWWWWATWDTWW 1029

Qy        251 ATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTGATTCCAGCGG 310
              ::: : ||| :|:| :|    |:: ::  ::    : ::: | :|:|  :::    : ::
Db       1028 DKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRRDGGRKRKDK 969

Qy        311 TGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGGATGTTTTGTT 370
              ::    ::::  |:       | :|  :::::::::::   |::  |::::|  :::::
Db        968 KDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDKDDD 909

Qy        371 ATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAG 430
               : :: ::: :| :::::::: :: :  :  :  | :   | :::::|: : : :| :|
Db        908 GKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTWDAA 849

Qy        431 AGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCGAATTTTATTA 490
              |::::|::: :| |:  :  : :   |::::: ::  :  ||:     ::|::: :  :
Db        848 ADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRDAAD 789

Qy        491 ATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGATGCTTTTAATT 550
              : : || :  :|||   :|::::::|: :   | ::  :::: : :::|     ::::|:
Db        788 DRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTW 729

Qy        551 TGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCAAAAGACTATC 610
              :::: ::  ::: |: ::|| ::   : :: |:   |  | :: :::: | :::   ::
Db        728 KRRWKRRDTRWDDADADDTARDDRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDAWWA 669

Qy        611 CGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAATTGTAATGATT 670
               |: || :|: ::   :::::: |   ::||     ::: : |: :| :: : : ::::
Db        668 DAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRADRWD 609

Qy        671 TTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTGTC 730
               :| :|:      :: : : :|::   : :  ||:|:: :  :  ||   :||      :
Db        608 RWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTTATWTTWAARAAWW 549

Qy        731 AATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAATTAA 790
              |::  | |  | |  :|  : |||| || ||     |:: || : ||:  : | |:  ||
Db        548 AWWAWATTTATWTTTWTTTWTTTTTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWAAAA 489
```

```
RESULT 4
BX145762/c
LOCUS       BX145762               781 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-108G13, genomic survey sequence.
ACCESSION   BX145762
VERSION     BX145762.1  GI:27977115
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 781)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 108G13. 108G13
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .781
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-108G13"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.9%;  Score 66.6;  DB 10;  Length 781;
  Best Local Similarity   46.3%;  Pred. No. 0.0023;
  Matches  219;  Conservative    0;  Mismatches  254;  Indels    0;  Gaps    0;

Qy         92 ATGATGTTAATATTATTTGTCTAACTGGTATATCTGAGGTAAAGCCAACACAAAATATTA 151
              || || |||||| ||    |   | |  || |  | | |||||    |   | |||| ||
Db        666 ATAATATTAATAATACAATTAATATTAATAAAAATAAAGTAAATAATAATAATAATAATA 607

Qy        152 ATATTCATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAG 211
              ||| | || ||   | |||  |||| ||| ||          |||   |  | || |
Db        606 ATAATAATAATAAAATTATTAATAATAATAATAACGTTAACAGTAGTTATAAAATTAATA 547

Qy        212 TAAAAAAAATAATTGTCGCCTTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTA 271
              || || |  |   |      | ||   | ||| | |  ||| | |||    | ||  ||
Db        546 ATAATAATAATAAAATTAATATTAATAAAAATAAAGTTAATAATAATAATAATAATAATA 487

Qy        272 ATATTTTTAGTCGTTTTATTAGGATGCTGATTCCAGCGGTGCCCCTGATATGTACCGCAC 331
              | ||| ||| |  |  || ||   |  | |          |    ||
Db        486 AAATTATTAATAATAATAATAACGTTAACAGTAGTTATAAAATTAATAATAATAATAATA 427

Qy        332 ACAACAAAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAG 391
              | || |  |||     |   |||  |   |    |  | ||      | | ||    |
Db        426 ATAAAATTAATATTAATAAAAATAAAGTTAATAATAATAATAATAATAATAATAAAATTA 367

Qy        392 CTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTA 451
              |  || || ||  | |  |||||  |      || |  |  |||      | ||    ||
Db        366 TTAATAATAATAACATTAACAGTAATTATAAAATTAATAATATTAATAATACAATTAATA 307

Qy        452 CACCTAAAAATAAAATAGTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTG 511
                | | |||| | |  |  |||  || |||||  |||| | || ||  |
Db        306 ATAATGATAATAATAAAGTAAATAATAATAATAATAATAATAATAATACAATTAATAATA 247

Qy        512 ATATTAATGTCAGAAAGAAAACGCGAGATGCTTTTAATTTGAAAGACAGTACA 564
              ||| ||||    ||| || |    |     |||  |||  | | ||||
Db        246 ATAATAATAAAGTAAATAATAATAATAAGAAGAAGAATAAGAAGAAGAATACA 194


RESULT 5
CNS020K7
LOCUS       CNS020K7               1092 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
```

222L11 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL175696
VERSION AL175696.1 GI:7813753
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
PUBMED 10899143
REFERENCE 3 (bases 1 to 1092)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .1092
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="222L11"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG222CF06LP1 end : T7"
ORIGIN

Query Match 5.9%; Score 66; DB 10; Length 1092;
Best Local Similarity 36.5%; Pred. No. 0.0031;
Matches 151; Conservative 60; Mismatches 203; Indels 0; Gaps 0;



RESULT 6
CNS00EPO/c
LOCUS CNS00EPO 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069493
VERSION AL069493.1 GI:4949636
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29P01"
/clone_lib="RPCI-98"
/note="end : T7"
ORIGIN

Query Match 5.7%; Score 63.6; DB 10; Length 1101;
Best Local Similarity 33.6%; Pred. No. 0.0097;
Matches 202; Conservative 99; Mismatches 297; Indels 4; Gaps 1;

```
Qy        391 GCTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCT 450
Db        850 TGTAAATAATDTAGWWATAAAATTAAAAARAAWTWTTTTWATATWAWAAATTRTTAAWAA 791

Qy        451 ACACCTAAAAATAAAATAGTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTT 510
Db        790 ARWDRTAAAAWAATA----TAATTRTTAAAAAAWTATTTTTTATAAATAATTTWAWTWTT 735

Qy        511 GATATTAATGTCAGAAAGAAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTA 570
Db        734 WWTTTTTWGATAAWWAWAAAAAAKAWTAKTRARATAAATTATWATWATTATATATWATTT 675

Qy        571 CTGCTCGCAGTAGGAAGACTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATA 630
Db        674 TTATTTTTATWAWTAADWATTTAATAAATATAWTAATTTATTAATAAAATAWATWAAAAA 615

Qy        631 AATCATTTGATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGC 690
Db        614 ATTTTARADATTAAATTAWAAWTTAWAATAAAWAATTAWTGAAAATTTTWWTTATATTAW 555

Qy        691 GATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAA 750
Db        554 TTWTTKTAAKTAAAATRTTAWATTAATTTATWTAWATTTTWWTTTTTWWTTTATTAAA 495

Qy        751 GT 752
Db        494 AT 493
```

RESULT 7
CNS0182P
LOCUS CNS0182P 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL108811
VERSION AL108811.1 GI:5629115
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37D10"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : SP6"
ORIGIN

Query Match 5.6%; Score 63.4; DB 10; Length 1101;
Best Local Similarity 21.4%; Pred. No. 0.011;
Matches 113; Conservative 208; Mismatches 204; Indels 2; Gaps 2;

```
Qy        330 ACACAACAAAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTT 389
Db        575 AAAAAADDTKDRWDARGAAARAAAAADAAARWAAKWTWAAAAARAAAAAAAAAAAAAAA 634

Qy        390 AGCTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGC 449
Db        635 AGTTKTTAAAAAAAGGGWGKTGKGKAADGDGGAAAADAWRRGRGRGRRARAARAAAAADA 694

Qy        450 TACACCTAAAAATAAAATAGTAGAGATTCCGAATT-TTATTAATACAAATAAATTTGATT 508
Db        695 KRAAAAAAAAAAAAAAAAAADAGGKRKWKKGDGKDGGKATKTAAWAKGRKDWGTATAWWT 754

Qy        509 TTGATATTAATGTCAGAAAGAAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAG 568
Db        755 WDATWADTWKAATDTDAKRAAAGRRKRDARKTARDGGRRARTRRRAWAAGGKRARAGARR 814

Qy        569 TACTGCTCGCAGTAGGAAGACTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAA 628
Db        815 ARAARRAADDRDWDAWAAAAAAAAAAAAAAWTTWRDRWWDWDDWDTRWDDDTTAAWWDDAR 874

Qy        629 TAAATCATTTGATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTG 688
Db        875 ARARRRRRRRRRRRRARRAARRAADDTDTKDRWADATTDKDTTKTWTTDDDDWDKAKRDR 934

Qy        689 GCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATA 748
Db        935 WWAAKADGAWKWRDRARDDWAATAKDDDGWKDKWGGRGKRGGKDKKRWDKGKTGKKDDDD 994

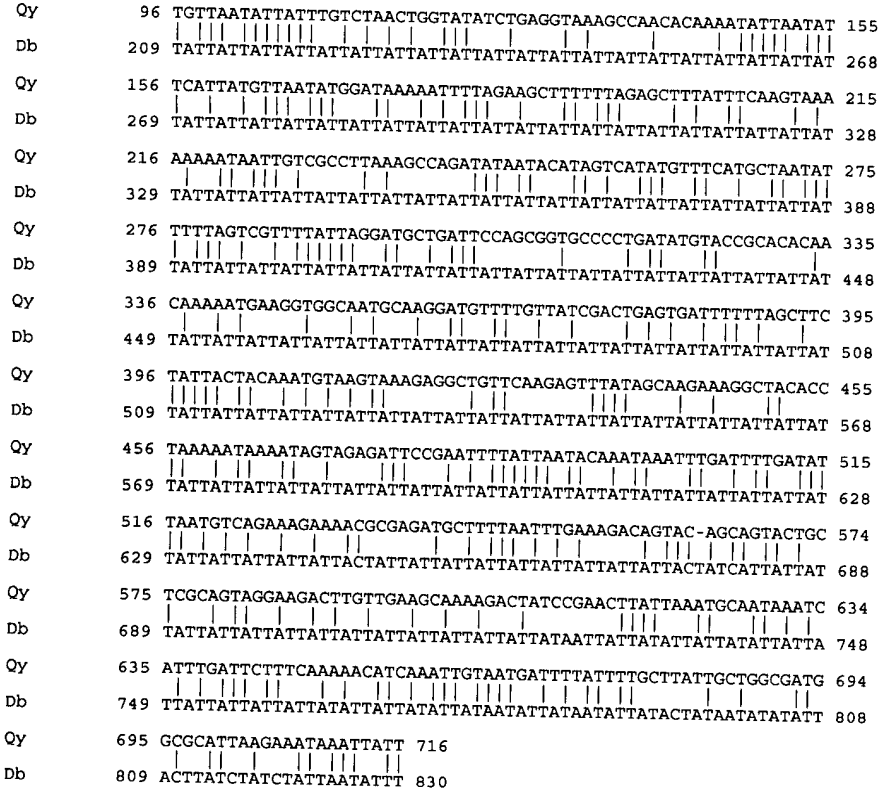
Qy        749 AAGTTTTCTTCTTGGGGCAAAGAAGTGATATTA-AAGAATTAATGTGTGCTGCAGATCTT 807
Db        995 WDKTWTRDWWWWWTRTKWDWWWDDGRGRGRWTRKRWGAWWRADAWARDDTDGKDTRTADK 1054

Qy        808 TTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCGTTGTTGCAGA 854
Db       1055 DRKRTDTTKRDGDDWRKDRKDKKRRDKKGDDKTKKDATWDDDDARDW 1101
```

RESULT 8
CW952279
LOCUS CW952279 891 bp DNA linear GSS 20-DEC-2004
DEFINITION TcB38.1_B02_T7 Tribolium BAC library Tribolium castaneum genomic, genomic survey sequence.
ACCESSION CW952279
VERSION CW952279.1 GI:56735316
KEYWORDS GSS.
SOURCE Tribolium castaneum (red flour beetle)
ORGANISM Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tribolium.
REFERENCE 1 (bases 1 to 891)
AUTHORS Savard,J. and Tautz,D.
TITLE Tribolium castaneum BAC-ends sequencing project
JOURNAL Unpublished (2003)
COMMENT Contact: Savard, J.
Abteilung fur Evolutionsgenetik, AG Tautz
Institut fur Genetik, Universitat zu Koln
Weyertal 121, 50931 Koln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: savard@uni-koeln.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .891
/organism="Tribolium castaneum"
/mol_type="genomic DNA"
/strain="GA-2"
/db_xref="taxon:7070"
/clone_lib="Tribolium BAC library"
/note="Vector: pBACe3.6; Site_1: EcoR1; Site_2: EcoR1; Library constructed by Exelixis Inc."
ORIGIN

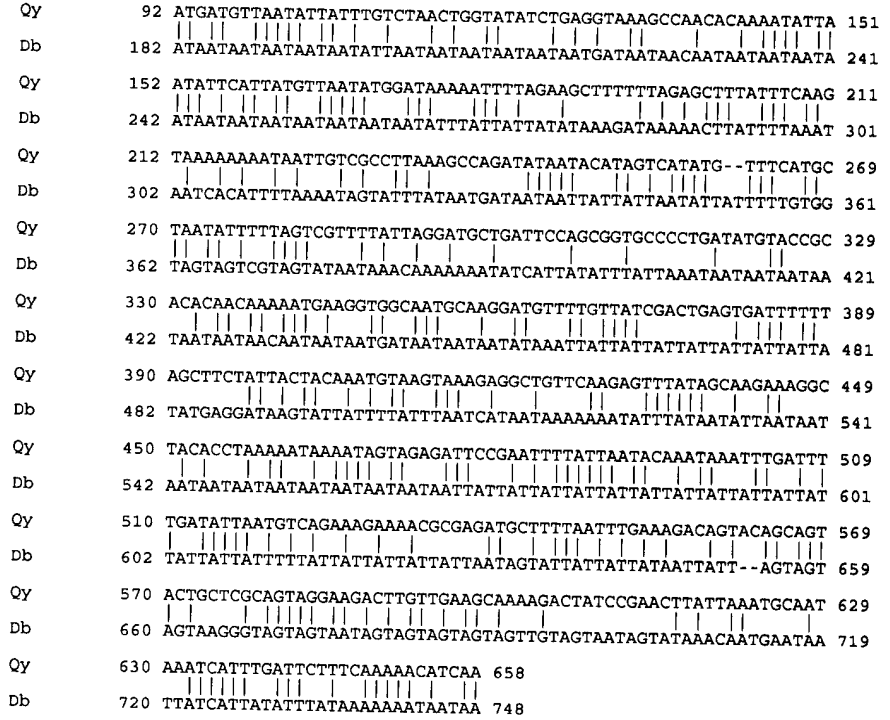
Query Match 5.4%; Score 61.2; DB 10; Length 891;

Best Local Similarity 44.7%; Pred. No. 0.03;
Matches 278; Conservative 0; Mismatches 343; Indels 1; Gaps 1;



RESULT 9
BX154986
LOCUS BX154986 755 bp DNA linear GSS 28-JAN-2003
DEFINITION Danio rerio genomic clone DKEY-114G23, genomic survey sequence.
ACCESSION BX154986
VERSION BX154986.1 GI:27986559
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 755)
AUTHORS Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Unpublished
COMMENT This sequence was generated from the T7 end of BAC 114G23. 114G23 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES Location/Qualifiers
source 1. .755
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-114G23"
/tissue_type="Testis"
/note="vector pIndigoBAC-536"
ORIGIN

Query Match 5.4%; Score 61; DB 10; Length 755;
Best Local Similarity 46.6%; Pred. No. 0.033;
Matches 265; Conservative 0; Mismatches 300; Indels 4; Gaps 2;



RESULT 10
CL501343/c
LOCUS CL501343 1055 bp DNA linear GSS 01-APR-2004
DEFINITION SAIL_696_B04.v1 SAIL Collection Arabidopsis thaliana genomic clone SAIL_696_B04.v1, genomic survey sequence.
ACCESSION CL501343
VERSION CL501343.1 GI:45998663
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1055)
AUTHORS Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D., Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D., Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B., Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
TITLE A high-throughput Arabidopsis reverse genetics system
JOURNAL Plant Cell 14 (12), 2985-2994 (2002)
PUBMED 12468722

```
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS830677; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .1055
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_696_B04.v1"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN

  Query Match             5.4%;  Score 60.6;  DB 10;  Length 1055;
  Best Local Similarity   35.9%;  Pred. No. 0.041;
  Matches  289;  Conservative    0;  Mismatches  512;  Indels    5;  Gaps    2;

Qy         48 ACACCTTATGACGCAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTAT 107
              |    ||||        | |  | |   |  |    || |           ||  | |    ||
Db        998 ANNTNTTATNTNNNNATNANTATANNATNTNNNTANATANATTANNAATAATTAATNAAT 939

Qy        108 TTGTCTAACTGGTATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAA 167
               | |    |     |||| |    |||    |  | || ||  || |||  | |||    |
Db        938 ATNTANTATNNTAATATATANTATAATNAAATNATAATATTATANTATNTAATATAAATA 879

Qy        168 TATGGATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGT 227
              |      |||    | || ||  || | |       || ||        ||| | ||||   |
Db        878 TTATATTAATNTATNTATAANNTTATNTTATNNATTTTTAANTTATAAATATATAANNNT 819

Qy        228 CGCCTTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTT 287
                | ||    | | |||| |   || |   ||    | |    | ||| | |||    | |
Db        818 AANNTAAATATTA-ANATAANATNTAATNTAATTAATAANATTTATAATATTATATATAT 760

Qy        288 TATTAGGATGCTGATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAA----AAATG 343
                |   ||    |    |          |  ||  ||    | |  | |    || |
Db        759 ATATTAAATATATAAAATATATTANAATATANTAAATANATAATAATAATATATTAANTT 700

Qy        344 AAGGTGGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTA 403
              ||      ||   |  ||  |  |||       |   ||  |  |      || | ||
Db        699 AAATAAATAANATATATATANTNANTTAANTTAATANATATANTNAATTAATAATAATTA 640

Qy        404 CAAATGTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATA 463
               |  |     || | |   | |  |  |  ||||   |   ||    | |  || |  ||
Db        639 NATNTNATTATATATAAAATATNTATTAANTTATTTAANNTAAATAAATATATATATNTA 580

Qy        464 AAATAGTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCA 523
              |  |  ||          |  | |||||  | | |  |   |  || || |||   || |
Db        579 ATTTTATAATATAATATATATATTATNATATAATATATATATTAATAAATATATAT 520

Qy        524 GAAAGAAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAG 583
                | | ||          | |  | |  ||  | |     |    || |
Db        519 ANNAAATAANAATNATAATTATATAATANAATAATATATAATATATAATATAATTTAATA 460

Qy        584 GAAGACTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTC 643
               |  | |  ||  |  | ||   |||    |  | ||  |||  |||| || | ||  ||
Db        459 TATAAATAATTATATTATAAATTTATTATATATAATATAATAATATATTAATTNTTTN 400

Qy        644 TTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAA 703
              ||          |||    |   || |||     ||  |     |
Db        399 NTTTNNNNNNTNNTNATTTNNNTNTNNTTTTTTNTNNNNTTTTTNNNTTTNNNNTNNNN 340

Qy        704 GAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGG 763
                  |   || ||    ||  | | |   |       ||             |  | ||
```

```
Db        339 NNNTTNNNTTNTTTNNNTTATTNTNTTNTTNTNNNTTTNNNNTNNTNNANNTTNTATTNT 280

Qy        764 GGCAAAGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTG 823
                      |    ||       ||    | |     | |  | | | | ||
Db        279 NNNNNTATTTTTNNNTTNTTTNNTTNNANTNTTNNTTANANTTNTATTTNTTNTTNNNNN 220

Qy        824 AGTGGGAAGGTTTTGGTCTCGTTGTT 849
                    ||||  | |   |  |
Db        219 NTNNTATNTATTTTNNTTTTNNTTNT 194


RESULT 11
CNS00EVL
LOCUS       CNS00EVL                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             5.3%;  Score 59.6;  DB 10;  Length 1101;
  Best Local Similarity   33.7%;  Pred. No. 0.065;
  Matches  196;  Conservative   88;  Mismatches  297;  Indels    1;  Gaps    1;

Qy        210 AGTAAAAAAAAATAATTGTCGCCTTAAAGCCAGATATAATACATAGTCATATGTTTCATGC 269
              | |: ::::::::| | |      :|||    | : ||:| | |:| : :  | ||: ::
Db        459 AHTWTMMMMMWWAATWTWWAAAWAAAWTTATWAATWAAAAAAWAWWWWATTTTTWWWWWT 518

Qy        270 TAATATTTTTAGTCGTTTTATTAGGATGCTGATTCCAGCGGTGCCCCTGATATGTACCGC 329
              :| |:|||: : |   :| |  |   |  | |||  |              | |:|  |
Db        519 WATTWTTTWAWWTWTAWTAAAAAAAAAWATAATTTAAAWWAATAWATTAAWAATTTAWAA 578

Qy        330 ACACAACAAAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTT 389
              :    |  || :|  |  |   | |   |  |      | || |  : | |    |||  |
Db        579 WWTATATTAATWTATAAATWTWATTAATATAAAAAAATATTTTTTWATAAAATTTTTAAT 638

Qy        390 AGCTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGC 449
```

```
Db        639 A-ATTTAATTAWTTATTAAATAAWTWATTWWWTAATTAAATAATTTWAAATAWAAAAAAA 697

Qy        450 TACACCTAAAAATAAAATAGTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTT 509

Db        698 AAAAAAAAWATWAAWAATWATAWATAAWTTAAAAWAATAAAAWAAWAATWAWWATAATAT 757

Qy        510 TGATATTAATGTCAGAAAGAAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGT 569

Db        758 WWATATATAWTTWTAWWWATWWAWWWTATAWATAWAATAWAAWAWAWATAAATAWATAWA 817

Qy        570 ACTGCTCGCAGTAGGAAGACTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAAT 629

Db        818 TWAAAWAWAWATAWWATWATATAWWAATAWAWAAAAAATWTAATATWAATWATAAWAAAA 877

Qy        630 AAATCATTTGATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGG 689

Db        878 AATAWAWTTWWTWTTTTWWAWWATATAAAWAWATAWAAAWAAAAAAAAAAAAATAAWAWW 937

Qy        690 CGATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAA 749

Db        938 WTWATATTWTTATTAAAWTWTATWWATTWATTWWAWTWTATAWTTWTATWTATATATWWT 997

Qy        750 AGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAATTAAT 791

Db        998 AWTAWATATATTTATTAAWWTATATTTAWAAAWTAATATAT 1039
```

RESULT 12
CC206942/c

```
LOCUS       CC206942               1563 bp    DNA     linear   GSS 09-MAY-2003
DEFINITION  CH261-191L18_Sp6.1 CH261 Gallus gallus genomic clone CH261-191L18,
            genomic survey sequence.
ACCESSION   CC206942
VERSION     CC206942.1  GI:30491401
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1563)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000  Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 301
            High quality sequence stop: 608.
FEATURES             Location/Qualifiers
     source          1. .1563
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-191L18"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN
```

```
Query Match           5.2%;  Score 58.6;  DB 9;  Length 1563;
Best Local Similarity  44.9%;  Pred. No. 0.11;
Matches  261;  Conservative    0;  Mismatches  319;  Indels    1;  Gaps    1;
```

```
Qy         98 TTAATATTATTTGTCTAACTGGTATATCTGAGGTAAAGCCAACACAAAATATTAATATTC 157

Db       1324 TTATTTTTATTTTATTGTATTAATTTATTGTGATTAATAATTTTTTAATAATATTTTATAT 1265

Qy        158 ATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAA 217

Db       1264 TTTTAATATTTATAATTTTTTATTTTAAATTTTATTTAGATTATAAATATTTTTTATTAA 1205

Qy        218 AAATAATTGTCGCCTTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTT 277

Db       1204 AATTTTTATTATAAATATTATTTTATTAATTATAATTTTTTATATTTTAATTTTATATTT 1145

Qy        278 TTAGTCGT-TTTATTAGGATGCTGATTCCAGCGGTGCCCCTGATATGTACCGCACACAAC 336

Db       1144 GTAATAATATATAATATATATTAATTTTTTATTAAAATTTTTTTTTTTTACAAATAAT 1085

Qy        337 AAAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCT 396

Db       1084 TATAAATAATTTATTATTTTTTTTTTATGAATTTAATTTTAAATTTAAATATTTTTATTTT 1025

Qy        397 ATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCT 456

Db       1024 TTTTATTTTAATATTAATATTTAAATTTTTTTATATAATATTATATAAAATTTTATTTGT 965

Qy        457 AAAAATAAAATAGTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATT 516

Db        964 ATAAATAATTATTTATTTTATTAAAAATTTATTTAAATAATTAAATTTTTATTTTTTATT 905

Qy        517 AATGTCAGAAAGAAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTC 576

Db        904 AAATTTTTTAAGAAATTTAAATTTTTTTTTAATAATTTAAAATTATAATAATATATTAGT 845

Qy        577 GCAGTAGGAAGACTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCAT 636

Db        844 TTAGTATTAAAATTTTTTTTATAATCATTAAATATAATTTAATTAATTTGAAATAAATAT 785

Qy        637 TTGATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTT 677

Db        784 ATTTTTTTTTAAATATTGAGTAATAATAAAGAATTTCTTGT 744
```

RESULT 13
CNS014PQ/c

```
LOCUS       CNS014PQ                987 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN12P22 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL104456
VERSION     AL104456.1  GI:5616067
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .987
                     /organism="Drosophila melanogaster"
```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN12P22"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             5.2%;  Score 58.4;  DB 10;  Length 987;
  Best Local Similarity   28.9%;  Pred. No. 0.12;
  Matches  142;  Conservative  106;  Mismatches  243;  Indels    0;  Gaps    0;

Qy        332 ACAACAAAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAG 391
              | :| ||:| |  :       :   ||  |  :::: ::::  : |   |  ||||||
Db        986 AAMATAAMACTCCMNMAAMWTMAAAAAAAAAAAWWWWWWWWWWWTTTTTTWTTTTTTTTT 927

Qy        392 CTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTA 451
               ||  ::|   : :::: ::  | :: :   | || :| : |||:     : |:    ||
Db        926 TTTAAWWTTTAWTWWWWWHWTTTYHWWWWTWTTTTWWATWWTTTWATTWTWWAWYCYWTA 867

Qy        452 CACCTAAAAATAAAATAGTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTG 511
                 |:::||:||  ||  : | |     |::  ::|  |:: : || ||||: : |:
Db        866 TTTWTWWWAAWAATTTATAWWAAAAAAAAAWWAAWMTNTAWWAWTATTAATTWWWATWAA 807

Qy        512 ATATTAATGTCAGAAAGAAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTAC 571
               :| |:|  | :  ||  ||     : |   |||:|||| | |          | :|
Db        806 TWAATWAATTAWTTAATTAATWAAWWAAAAATTTWAATTTAATAAWTTTCTTTAAAWWAT 747

Qy        572 TGCTCGCAGTAGGAAGACTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAA 631
               :     : |      : || :| ||  |||| | :|:   ||  :   : ::  |:| |
Db        746 WWTAAWAWTTTWTTTWWTTTTWTTAAAAAAAAAAAWAWWWWAAWAWNANWTWWATAWTTA 687

Qy        632 ATCATTTGATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCG 691
               |  |||  || |:|     |    || || :  :  ::|| ||:|: |: : |  || |
Db        686 TTTTTTTTTTTWTWTTTTTTATTCTAATTTTWGTWWTKWTTKTTKTKKTKKKGGGGGGGG 627

Qy        692 ATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAG 751
               | :  ::  : ||| ||:  :    :   :  : ::  ::   : |: |   |||:
Db        626 GGGMVAGWKRTMAAAAAAAAMMAMMMMMMMNVMMVYYYYYYYHYTTYHTTWTTTTCTAAWH 567

Qy        752 TTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTG 811
               :: | :| :   :|:  |    | |:  | | ::||: |||    |     |::||::
Db        566 MHYCCMYCAWCYATHAWCTAKCACAMAYACACGCMWTARAGTGAATTTTCTTTMWTTHWT 507

Qy        812 TTTTGAGTTCT 822
               ::: : || |
Db        506 CYYHYHTTTTT 496


RESULT 14
CNS0039G
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.2%;  Score 58.4;  DB 10;  Length 1101;
  Best Local Similarity   20.4%;  Pred. No. 0.12;
  Matches  159;  Conservative  288;  Mismatches  327;  Indels    6;  Gaps    3;

Qy         18 AATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACGCAGTTAGCAGACCAAAT 77
              :: :|  |||   :   | ||  ::: :       || | ||  :| ||   :
Db        322 RRRRAAAGGGRRRRRRAGGGGRARRRRRRRRRRRGGGRGARGGAGGRRGGAGGGGRGGR 381

Qy         78 GTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGTATATCTGAGGTAAAGCC 137
              |     : | |||:||      :::::|| ||| | :|| :   |||  | :  :|:|
Db        382 GGGGGGRGGGGGGMATATAWAWWWWTTTTTTTTAWAAAWAAAATAATTWWAAWAWAAAA 441

Qy        138 AACACAAAATATTAATATTCATTATGTTAATATGGATAAAA--ATTTTAGAAGCTTTTTT 195
              ||   :||| :  ||:|:| ::| | :|:| |   | ||||  :||||   :  | :||:
Db        442 AATTWWAAAAWAAAAWAWTAWWTTTAWTWAAAAAAAAAAATTWTTTTTTTWTTTAWTTW 501

Qy        196 AGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCAGATATAATACATAGT 255
              | |  ||:::|| ||  |||||||| |:   :     ||||   :  ::|  |  | : |
Db        502 ATAWWTTWWWTTAAAWAAAAAAAAAAAWAAAWAAAWATAAATWTWWTWWTTYTTWAAWAT 561

Qy        256 CATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTGATTCCAGCGGTGCCC 315
               | :    :: : :|  | :::|:: :  |::::|    |: :| ::: :: :  :  :
Db        562 AAAMCMAAWYYHTYTTYTYHYYTYWTYTMTWHYHTMYTHAWAHTTWYHWYHTYAMWHWMT 621

Qy        316 CTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCGA 375
              :   :  ::   :  |:  :   : : :   :   :|:  |:   : ::|: ::: :
Db        622 WHWTMWAWWHWTTYTAAYYYYYTCMYYYHYMHWHHAHAHAAAWWTTHTWWTHAYHWATYH 681

Qy        376 CTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTTT 435
              ::   : |     |: |::|: :::::  :: : ::: :   :  :: :: :    |:
Db        682 YYYYMYCAMMCMCTHTCHHCYYYYHHYTAHHTHTHHWYAHYYMWYYMWAYYWMYCTACTY 741

Qy        436 ATAGCAAGAAAGGCTACACCTAAAAATAA--AATAGTAGAGATTCCGAATTTTATTAATA 493
              ::: ::: ::|   |::|  :|:::::::|  || | :| :  |::: : |:: ::| :::
Db        742 HYHHHHHYHWAYHTTWYAWAHAMWMWHHAHYAAAAAWAAWATTHHYHHTTHYMHHTYMYH 801

Qy        494 CAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGATGCTTTTAATTTGA 553
              : : |     | : ::: :||:| :|  |:: ::   :: :   : :|  :: :::|:|
Db        802 YYMYTCCYMCTYHCWHYYHTAYTCWTWTHHWMWTWTHWYHHTWWHHTTTHWAWWHTHTWC 861

Qy        554 AAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCAAAAGACTATCCGA 613
              :: ::| |: :  : :||  ::  :: ::  :: ||:  :: :: : ::: :|| :::
Db        862 WWWWHATTWTWATHCWACMTMHWHHWMHMHHHHMACHAHHHTHMCMCHHHMHCTCHHHHT 921

Qy        614 ACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAATTGTAATGATTTTA 673
              :::   ::::: :::::::::||:  :||:|| :::  ::::: :::  :::    ::
Db        922 MYHMTCHWMWHMHWHWHHWWMATW--MTTMTTMMMMCCMMHHHCHMYHMMHMYMYCCHYY 979

Qy        674 TTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTGTCAAT 733
               | : | :: :   | ::      :||  :|:||:| |  |  :: :   :: | :::
Db        980 CTCHTHATTHYHYMCTCYHYCTWHTYWTAYWWAWTAHAMTTATWWWWMHWWAHWATWWWW 1039
```

```
Qy        734 TGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAATTAATGT 793
              : ::   :      :::  ::|:||::::     :::| :: : ::: :|:  :|:: :: |
Db       1040 WWWWATAWWACTCHHWTWYHTHCTWYYHHTYHMWWAWWMAWHWHHMYAHYHWAHHCWYYT 1099


RESULT 15
CNS0100X
LOCUS       CNS0100X                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098379
VERSION     AL098379.1  GI:5609990
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN03G04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             5.2%;  Score 58.2;  DB 10;  Length 1101;
  Best Local Similarity   18.1%;  Pred. No. 0.13;
  Matches  139;  Conservative  307;  Mismatches  314;  Indels    6;  Gaps    3;

Qy        243 TATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTGAT 302
              | :|: : |||:: ::    |  |   :||  |:|:|  | ||:|: || : :|: |
Db        335 TBDADGWWATARWAWDCRYATATAAAADAACTTWTKTTTTTGTDTDKTTTKTDTKTTTTR 394

Qy        303 TCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGGAT 362
              |   | |  :        |  |:   : :    | : : :|::  :|:   : : : : :|
Db        395 TTKGGGGTGKRTGNGTGTGAGKGDGKGDGGKGAKWTDTRTDDTWKTDTDTWGKADTDWWT 454

Qy        363 GTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAGGC 422
              :|::| :|:|    ||::: :|:|:|::: :: | :|: :  :  ::: ::|: ::: :
Db        455 DTWKTTWTWTWTTTTGWKKWDTKTKTKWKDDDGTTWTDTWKDWKTDKDTDKTWTDVWADN 514

Qy        423 TGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCGAA 482
              |:: : :  : : ::   :    :  |      |:: ::: : : :|:::|::|:  :|:
Db        515 AGWWTDKDKTKAKKKRTKTDTKKTKTRTGTKKKTDRGMRKTDTWGWGDWRADRTDGARAD 574

Qy        483 TTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGATGC 542
                |:: : :  |:|:| :|:|: |||:|: |:: : : ::: : :: : : :||:::
Db        575 RAANAKDKDADGKADTDATKTDGWGTTTKADTTDDTDADWWDWKDWDWKDADTKAGRKRG 634

Qy        543 TTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCAAA 602
              : :    ::   :   ::      ||: :  : : : : ::: ::   : :: ::
Db        635 WGWGKKGGWKKTTKTKKKKTDKTKTGTDTKWGWKDKWTDKKRDGKR--GDGDRKDKGTGG 692

Qy        603 AGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAATTG 662
              :|:::|:  :|: : |:  :|   :|||||| |: | : | |||        |    ||
Db        693 DGDSKAKGDRAWAWRAKATRAAAADATAAATAAKATAWATKTTTKKTTTTTTTKTTTTTT 752

Qy        663 TAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTT 722
              |  |  :|||| : ::  ::::: ::: : :|: : |:::::| ::::::|:::|  :
Db        753 TTTTTKKTTTARKADR--DDDDKAKRDGDKDGDWDDADKDWKAGRDRDDDTRKWGWKDRA 810

Qy        723 GGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAA 782
              |:::: :  ::: |::  ::|| |:| : :| | :| : :  :  :|: | |: :   |
Db        811 GKKKKAKKKDHKTAWWWADDTGTAWATWAKTATAKTGKAKATKDBKRGRGGGGRGDGRGA 870

Qy        783 AGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCT 842
               ::   :  : | : :: ::::  |:: |: :|||||:  :||::::|:|: : :::: :
Db        871 GRRKRARGRKRGRRRKDKRKRKKRTKKAGDGDTGAGTD--KAGDKKKADGDGWGWKRKWD 928

Qy        843 CGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAGT 902
              ::|::||    ::  |::|: : | |:|       | :|: |  |  ::::    : :: :
Db        929 SKTWKTTMTGTDDTTTWKGRNRRGRGRAGAGRRRGRKGKAGKGAGGKWWKWAWADABWTW 988

Qy        903 TAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTGGC 962
              |::::::::  |: |:    : :  :   ::       :  :| :   : ::::::
Db        989 TWDWKDWKWKWTDWGWGKGGWGDGRKGGAKDGAKRNRKNRDGMRYGNRAWKWMDDAATAK 1048

Qy        963 AGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAA 1008
              |:: ::::: ::::: : ::::: | ::: : :||     :: :::
Db       1049 ARWADDDWDADYDDDADADYWKWKATWMDBWCDTAKWTANWKNDWW 1094


Search completed: May 11, 2006, 05:00:56
Job time : 3606.15 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 20:34:42 ; Search time 156.922 Seconds
(without alignments)
12743.667 Million cell updates/sec

Title: US-09-423-093B-1_COPY_11821_12945
Perfect score: 1125
Sequence: 1 atgaagattgtttatataat.....gtaataatataattgattga 1125

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	245.8	21.8	4604	3	US-09-830-807-22	Sequence 22, Appl
	2	201.8	17.9	1077	3	US-09-328-352-1480	Sequence 1480, Ap
c	3	65.6	5.8	19250	3	US-08-961-527-35	Sequence 35, Appl
c	4	58	5.2	1830121	3	US-09-557-884-1	Sequence 1, Appli
c	5	58	5.2	1830121	3	US-09-643-990A-1	Sequence 1, Appli
c	6	58	5.2	1830121	3	US-10-158-865-1	Sequence 1, Appli
	7	57.2	5.1	1095	3	US-09-489-039A-933	Sequence 933, App
c	8	56.8	5.0	6474	3	US-08-961-527-155	Sequence 155, App
	9	55.2	4.9	525	3	US-09-583-110-2189	Sequence 2189, Ap
	10	55.2	4.9	660	3	US-09-107-433-1123	Sequence 1123, Ap
	11	54.6	4.9	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	12	53.4	4.7	1664976	3	US-08-916-421B-1	Sequence 1, Appli
c	13	53.4	4.7	1664976	3	US-09-692-570-1	Sequence 1, Appli
	14	52.8	4.7	1664976	3	US-08-916-421B-1	Sequence 1, Appli
	15	52.8	4.7	1664976	3	US-09-692-570-1	Sequence 1, Appli
	16	52	4.6	14602	2	US-08-597-236-1	Sequence 1, Appli
	17	52	4.6	14602	2	US-08-746-682A-1	Sequence 1, Appli
	18	50.8	4.5	9326	3	US-10-138-701-51	Sequence 51, Appl
c	19	50.8	4.5	17310	3	US-08-956-171E-23	Sequence 23, Appl
c	20	50.8	4.5	17310	3	US-08-781-986A-23	Sequence 23, Appl
	21	50.6	4.5	1044	3	US-10-138-701-56	Sequence 56, Appl
c	22	49	4.4	615	3	US-08-998-416-186	Sequence 186, App
c	23	49	4.4	636	3	US-08-998-416-1137	Sequence 1137, Ap
c	24	49	4.4	837	3	US-08-998-416-288	Sequence 288, App
	25	48.8	4.3	832	3	US-09-621-976-2813	Sequence 2813, Ap
c	26	48.4	4.3	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	27	48.2	4.3	50000	3	US-09-662-254B-23	Sequence 23, Appl
	28	48.2	4.3	50000	3	US-09-662-254B-24	Sequence 24, Appl
	29	48.2	4.3	640681	3	US-09-790-988-1	Sequence 1, Appli
	30	47.4	4.2	1182	3	US-09-601-198-18	Sequence 18, Appl
	31	46.8	4.2	1086	2	US-08-415-751-47	Sequence 47, Appl
	32	46.4	4.1	225127	3	US-09-949-016-16480	Sequence 16480, A
c	33	46.2	4.1	50000	3	US-09-662-254B-24	Sequence 24, Appl
	34	45.8	4.1	1055	3	US-09-806-708B-23	Sequence 23, Appl
	35	45.6	4.1	606	3	US-09-601-198-166	Sequence 166, App
	36	45.6	4.1	1830	3	US-09-662-254B-67	Sequence 67, Appl
c	37	45.6	4.1	32392	3	US-09-662-254B-27	Sequence 27, Appl
	38	45.4	4.0	1155	3	US-09-891-641-37	Sequence 37, Appl
c	39	45.2	4.0	13321	3	US-08-956-171E-4	Sequence 4, Appli
c	40	45.2	4.0	13321	3	US-08-781-986A-4	Sequence 4, Appli
	41	45	4.0	1062	3	US-09-328-352-1104	Sequence 1104, Ap
	42	45	4.0	50000	3	US-09-662-254B-26	Sequence 26, Appl
c	43	44.8	4.0	1786	3	US-08-956-171E-618	Sequence 618, App
c	44	44.8	4.0	1786	3	US-08-781-986A-618	Sequence 618, App
	45	44.8	4.0	76124	3	US-09-949-016-15976	Sequence 15976, A

ALIGNMENTS

RESULT 1
US-09-830-807-22
; Sequence 22, Application US/09830807
; Patent No. 6846667
; GENERAL INFORMATION:
; APPLICANT: Crooke,Helen R.
; APPLICANT: Clarke, Enda E.
; APPLICANT: Everest, Paul H.
; APPLICANT: Dougan, Gordon
; APPLICANT: Holden, David W.
; APPLICANT: Shea, Jacqueline E.
; APPLICANT: Feldman, Robert G.
; TITLE OF INVENTION: VIRULENCE GENES AND PROTEINS, AND THEIR USE
; FILE REFERENCE: GJE-65
; CURRENT APPLICATION NUMBER: US/09/830,807
; CURRENT FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 4604
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2982)..(4082)
; NAME/KEY: CDS
; LOCATION: (1534)..(2637)
; NAME/KEY: CDS
; LOCATION: (749)..(1531)
; NAME/KEY: CDS
; LOCATION: (6)..(746)
US-09-830-807-22

Query Match 21.8%; Score 245.8; DB 3; Length 4604;
Best Local Similarity 55.8%; Pred. No. 6.3e-48;
Matches 550; Conservative 0; Mismatches 402; Indels 33; Gaps 3;

```
Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||| ||  | | | | || || ||  |      ||| || || ||  | |   |
Db       1534 ATGAAAATACTATTTGTCATTACAGGTTTAGGCCTTGGAGGTGCTGAGAAGCAGGTTTGT 1593

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
              |  ||||| ||  || |   | || |||||||  |||| || |||||||  ||    |
Db       1594 CTTTTAGCTGATAAATTAAGTTTAAGCGGGCACCATGTAAAGATTATTTCACTTGGACAT 1653

Qy        121 ATATCTGAGGTAAAGCCAACAC---------AAAATATTAATATTCATTATGTTAATATG 171
               || ||| |    ||       |         | ||| ||||| |   | |||| ||||||
```

```
Db       1654 ATGTCTAATAATAAAGTCTTTCCTAGCGAAAATAATGTTAATGTCATTAATGTAAATATG 1713
Qy        172 GATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCC 231
Db       1714 TCAAAAAACATTTCTGGAGTTATAAAAGGTTGTGTCAGAATTAGAGATGTTATAGCTAAT 1773
Qy        232 TTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATT 291
Db       1774 TTCAAACCAGACATTGTACACAGTCATATGTTTCATGCAAACATTATCACTAGATTGTCT 1833
Qy        292 AGGATGCTGATTCCAGCGGTGCC---CCTGATATGTACCGCACACAACAAAAATGAAGGT 348
Db       1834 GTAATTGGAATCAAAAACAGACCTGGTATTATATCAACTGCACATAATAAAAATGAAGGT 1893
Qy        349 GGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAAT 408
Db       1894 GGGTATTTCAGAATGCTCACATATAGAATAACCGATTGTTTAAGTGATTGTTGTACAAAT 1953
Qy        409 GTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATA 468
Db       1954 GTTAGCAAAGAAGCAGTGGATGAGTTTTTACGGATAAAAGCCTTTAATCCCGCTAAAGCA 2013
Qy        469 GTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAG 528
Db       2014 ATTACTATGTATAATGGGATAGATACCAATAAATTTAAATTTGATTTATTGGCAAGGAGG 2073
Qy        529 AAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGA 588
Db       2074 GAAATTCGAGACGGTATTAATATAAAAAATGATGATATATTATTACTTGCTGCAGGTCGT 2133
Qy        589 CTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCA 648
Db       2134 TTAACGTTAGCTAAAGATTATCCTAATTTATTGAATGCAATGACTCTGCT---------- 2183
Qy        649 AAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAAT 708
Db       2184 ----------TCCTGAACACTTTAAACTTATTATTATTGGTGATGGTGAATTGCGTGAC 2232
Qy        709 AAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAA 768
Db       2233 GAAATTAATATGCTTATAAAAAAATTGCAATTATCTAATAGGGTGTCCTTGTTGGGAGTT 2292
Qy        769 AGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGG 828
Db       2293 AAAAAAAATATTGCTCCCTATTTTTCTGCATGTGATATTTTTGTTCTCTCTTCTCGTTGG 2352
Qy        829 GAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACC 888
Db       2353 GAAGGATTTGGATTAGTCGTGGCAGAAGCTATGTCATGTGAGCGAATTGTTGTTGGCACG 2412
Qy        889 GATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAAT 948
Db       2413 GATTCAGGGGGAGTAAGAGAAGTTATTGGTGACGATGATTTTCTTGTACCCATATCTGAT 2472
Qy        949 CATATTCTGTTGGCAGAGAAAATCG 973
Db       2473 TCAACACAACTTGCAAGCAAAATTG 2497
```

RESULT 2
US-09-328-352-1480
; Sequence 1480, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1480
; LENGTH: 1077
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1480

Query Match 17.9%; Score 201.8; DB 3; Length 1077;
Best Local Similarity 52.8%; Pred. No. 8.3e-38;
Matches 493; Conservative 0; Mismatches 422; Indels 18; Gaps 2;

```
Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
Db         13 ATGAAAGTTATGTTGTTTATTACTGGTTTAGGCATGGGAGGAGCCGAGAAAGTTGTTTGC 72
Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
Db         73 GAATTAGCTGATAAATTATATGAGTTGGGACATGAAGTCTGTATTGTATATTTTACGGG- 131
Qy        121 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 180
Db        132 -----TGAAATTATAAGACGACCGAAGAATAATATCAATATAATTTATATACCACTAAAA 186
Qy        181 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 240
Db        187 ATTAGTTCACTTATTTTTAGTGTATATTCAGTTATGAAACTAGTTGATCAATGGAAACCT 246
Qy        241 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 300
Db        247 GATGTAATCCATTCACATATGTTTCATGCAAATATAATTGCAAGATTGGCAAAAATAAGA 306
Qy        301 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 360
Db        307 AGAAAAAATATTCGTATAATTTGTAGTTCGCATTCTAATTATGAAGGTGGAGCATTACGT 366
Qy        361 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 420
Db        367 ATGATGATGTATGCTATGACTGAAAAATTATGTGATGTACATACAAATGTAAGTTCTAAT 426
Qy        421 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 480
Db        427 GCTGCTACTGCATTAATGCATGCTGGAGCAGTAAGAAAAAGAAAAATTAAAGCTGTTTAT 486
Qy        481 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540
Db        487 AATGGTATTAATATAAAAGACTATCATTTTGATGAAGTAGAAAACAAGTACTTACATAAA 546
Qy        541 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 600
Db        547 GCGTTTAACTTGAAAAATGAAGCTAAAACTATACTTGCAATAGGAAGGATAGATACTCCT 606
Qy        601 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 660
Db        607 AAAGATTATCCAAATTTATTAAAAGCTTTCAAAATAATTTATGAGATAGAC--------- 657
Qy        661 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 720
Db        658 ---AATAGTGTTCATCTATTTATAGCTGGAAATGGGCCAAAAAAAGAAGAAATCATGCTA 714
Qy        721 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 780
Db        715 TTGGCTAATAATTTAGGTATTAATAGCAATGTTAAATTTTTAGGCATACGGAATGATATT 774
Qy        781 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 840
Db        775 CCTAAGCTATTAAATTCATGTGATTTATTTATTTCATCGTCTGCATGGGAAGGATTTGGA 834
Qy        841 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 900
Db        835 CTAGCTATATTAGAGGCGATGATTTGTAAAAAACCTATTGTAGCAACTGAAACTGATGGT 894
Qy        901 GTTAAAGAAGTCGTTGGACCTCATAATGATGTT 933
Db        895 GCCAAAGAATTATTAAATAATGAAAATTTGGTT 927
```

```
RESULT 3
US-08-961-527-35/c
; Sequence 35, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19250 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-35

  Query Match             5.8%;  Score 65.6;  DB 3;  Length 19250;
  Best Local Similarity   47.8%;  Pred. No. 1.5e-05;
  Matches  265;  Conservative    0;  Mismatches  274;  Indels   15;  Gaps    2;

Qy        416 AAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGA 475
              ||||  | | | |   |||    | |   || ||        |    ||||| | |   |
Db      11753 AAGATTCGGCTAAGTGGTTATATGGAGAGAAAGCGTATAAAGACGGAAAAATTGAAATTA 11694

Qy        476 TTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGC 535
              ||   |||   ||| ||   || || | | | |||       |||| ||| | |||   |
Db      11693 TTTTTAATGCAATTGATTTGAAAAAGTATCAATTTTTGTCAGATGTTAGAGAAAAATGTC 11634

Qy        536 GAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTG 595
              |    |  ||  || ||  |   | ||     ||| |         |||   | ||   |
Db      11633 GTAGAGAATTAGATGTGTCA---AATAAGTTCGTATTAGGAAATATAGCTCGCCTATCAG 11577

Qy        596 AAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACAT 655
              |   |||| || |     | ||||| || |   ||||  |  | ||| |    ||| ||
Db      11576 ATCAAAAAAACCAAAGTTATTTATTTAACGTTTTAAAAGAACTCATTTTAATCAAACCA- 11518

Qy        656 CAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTAT 715
                  || ||||||| ||  | | |||  |||| |          | ||||||
Db      11517 ----------AATGTTATTTTACTCCTAGTTGGTAATGGTGAGGATGAGCAGAAATTAA 11469

Qy        716 TGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTG 775
                |    | ||   || | |||||         | ||  | || || |||  ||| | ||
Db      11468 AACAGAAAGCTTTAGAACTAAATCTGACCCCATATGTGCTATTTTTAGGGAGAAGGACTG 11409

Qy        776 ATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTT 835
              |||||   || ||| | | |||    ||| |||||  |  |    ||| | |  || || |
Db      11408 ATATTTCTGATTTATTATCTGCGATGGATGTTTTTTTGCTTCCGTCTAAATATGAGGGGT 11349

Qy        836 TTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTG 895
              |   | |  | | | ||| |||  ||| |  | |   |    || |  |  | ||| | |
Db      11348 TGCCTGTTTCTCTAGTAGAGGCTCAGGCATCGGGATTACAAATTTTATCGTCAGATACAG 11289

Qy        896 GTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTC 955
                  ||| ||  ||        | | |  |    | ||| |||   |  |   |
Db      11288 TGACGCAAGAAGTAGATGTGACCAAAAACATTAGTTACTTACCTATCAACGAAGAGTCTG 11229

Qy        956 TGTTGGCAGAGAAA 969
              |||||  |  ||||
Db      11228 TGTTGCTATGGAAA 11215


RESULT 4
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
;  GENERAL INFORMATION:
;       APPLICANT: Fleischmann et al.
;       TITLE OF INVENTION: The Nucleotide sequence of
;                           the Haemophilus influenzae Rd Genome, Fragments
;                           Thereof, and Uses Thereof
;       NUMBER OF SEQUENCES: 1
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3 1/2 inch diskette
;            COMPUTER: Dell Pentium
;            OPERATING SYSTEM: MS DOS v6.22
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/557,884
;            FILING DATE: 25-Apr-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/476,102
;            FILING DATE: JUN-5-1995
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Michelle S. Marks
;            REGISTRATION NUMBER: 41,971
;            REFERENCE/DOCKET NUMBER: PB186P3
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 301-309-8504
;            TELEFAX: 301-309-8439
;  INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1830121 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

  Query Match             5.2%;  Score 58;  DB 3;  Length 1830121;
  Best Local Similarity   47.7%;  Pred. No. 0.0033;
  Matches  226;  Conservative    0;  Mismatches  244;  Indels    4;  Gaps    2;

Qy        653 CATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAAT 712
              ||  ||| | |   ||||  | |||    || |  ||   |||||    |  | | || |
Db    1767381 CAAAAAAATATCCAGATTGGAATTTAAAAATAGTGGGTTCTGGCGAGGAAGAAGAAAATT 1767322

Qy        713 TATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAA 772
              ||    || | |    |||| || |||  |||  ||||   ||  |    | |  |
Db    1767321 TAAAAAATCTTGCTAAAGCATTGGATATTGAAGATTCAGTTAATTTTATTCCTCGTACTA 1767262

Qy        773 GTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAG 832
                |||| ||      |                  ||| |  |||   |||  |    ||||
```

```
Db    1767261 ATGATGTTTCTTTTTATTATGAAAGCAGTTCAATTTATTGTTTACCTTCACAAACAGAAG 1767202

Qy        833 GTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGC---TACCG 889

Db    1767201 GCTTGCCTTTAGTCGTTATTGAAGCGATGGCATTTGGCTTACCTATTGTTGCCTTTAATT 1767142

Qy        890 ATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGAT-GTTATCCCTGTCAGTAAT 948

Db    1767141 GTTCTCCTGGCGTAAAGCAATTAGTGGAACATAAAGAAAATGGTTTTCTTTGTGAACAAA 1767082

Qy        949 CATATTCTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAA 1008

Db    1767081 ATAATATTGAAGAAATGGTTAAAGGTTTAGATTTATTAATCAATAATCCTGANCTTTATC 1767022

Qy       1009 ATAATAGGTATGAAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTG 1068

Db    1767021 TACAAATGTCAGACAAATCTCGTTTAATGAGTGANGATTATGGCATTGAGAAAATTATTG 1766962

Qy       1069 AGTGAGTGGGAGCGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGAT 1122

Db    1766961 AAGAATGGAAAGGGATTTTATAAGGNATAATTTTATGTTAAAAAAATATTTGNT 1766908
```

RESULT 5
US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 5.2%; Score 58; DB 3; Length 1830121;
Best Local Similarity 47.7%; Pred. No. 0.0033;
Matches 226; Conservative 0; Mismatches 244; Indels 4; Gaps 2;

```
Qy        653 CATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAAT 712

Db    1767381 CAAAAAAATATCCAGATTGGAATTTAAAAATAGTGGGTTCTGGCGAGGAAGAAGAAAATT 1767322

Qy        713 TATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAA 772

Db    1767321 TAAAAAATCTTGCTAAAGCATTGGATATTGAAGATTCAGTTAATTTTATTCCTCGTACTA 1767262

Qy        773 GTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAG 832

Db    1767261 ATGATGTTTCTTTTTATTATGAAAGCAGTTCAATTTATTGTTTACCTTCACAAACAGAAG 1767202

Qy        833 GTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGC---TACCG 889

Db    1767201 GCTTGCCTTTAGTCGTTATTGAAGCGATGGCATTTGGCTTACCTATTGTTGCCTTTAATT 1767142

Qy        890 ATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGAT-GTTATCCCTGTCAGTAAT 948

Db    1767141 GTTCTCCTGGCGTAAAGCAATTAGTGGAACATAAAGAAAATGGTTTTCTTTGTGAACAAA 1767082

Qy        949 CATATTCTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAA 1008

Db    1767081 ATAATATTGAAGAAATGGTTAAAGGTTTAGATTTATTAATCAATAATCCTGANCTTTATC 1767022

Qy       1009 ATAATAGGTATGAAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTG 1068

Db    1767021 TACAAATGTCAGACAAATCTCGTTTAATGAGTGANGATTATGGCATTGAGAAAATTATTG 1766962

Qy       1069 AGTGAGTGGGAGCGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGAT 1122

Db    1766961 AAGAATGGAAAGGGATTTTATAAGGNATAATTTTATGTTAAAAAAATATTTGNT 1766908
```

RESULT 6
US-10-158-865-1/c
; Sequence 1, Application US/10158865
; Patent No. 6846651
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag
; Patent No. 6846651
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P2C1D1
; CURRENT APPLICATION NUMBER: US/10/158,865
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 09/557,884
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 08/476,102
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)

; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a,t,c, or g

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:

```
;   NAME/KEY: misc_feature
;   LOCATION: (145171)..(145171)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (145942)..(145942)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (147197)..(147197)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (150841)..(150841)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152500)..(152500)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature

  Query Match             5.2%;  Score 58;  DB 3;  Length 1830121;
  Best Local Similarity   47.7%;  Pred. No. 0.0033;
  Matches  226;  Conservative    0;  Mismatches  244;  Indels    4;  Gaps    2;

Qy        653 CATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAAT 712
              ||  ||| | |   |||| | |||   || | ||  |||||   | | | || |
Db    1767381 CAAAAAAATATCCAGATTGGAATTTAAAAATAGTGGGTTCTGGCGAGGAAGAAGAAAATT 1767322

Qy        713 TATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAA 772
              ||    || | | |    |||| || |||  ||| ||||   ||  |    | | |
Db    1767321 TAAAAAATCTTGCTAAAGCATTGGATATTGAAGATTCAGTTAATTTTATTCCTCGTACTA 1767262

Qy        773 GTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAG 832
               |||| ||     |                  ||| |  |||   |||  |    ||||
Db    1767261 ATGATGTTTCTTTTTATTATGAAAGCAGTTCAATTTATTGTTTACCTTCACAAACAGAAG 1767202

Qy        833 GTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGC---TACCG 889
              | ||   | | || |||   ||||| ||||| | ||      ||  |||||||    ||
Db    1767201 GCTTGCCTTTAGTCGTTATTGAAGCGATGGCATTTGGCTTACCTATTGTTGCCTTTAATT 1767142

Qy        890 ATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGAT-GTTATCCCTGTCAGTAAT 948
              |||||  ||| || ||  || | || | || | |   |  || ||| ||  ||| | ||
Db    1767141 GTTCTCCTGGCGTAAAGCAATTAGTGGAACATAAAGAAAATGGTTTTCTTTGTGAACAAA 1767082

Qy        949 CATATTCTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAA 1008
              ||  ||  |  |  |  ||  | | |||     ||| |                    |
Db    1767081 ATAATATTGAAGAAATGGTTAAAGGTTTAGATTTATTAATCAATAATCCTGANCTTTATC 1767022

Qy       1009 ATAATAGGTATGAAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTG 1068
                | | ||  || ||||   |  ||    ||   | | |  || | | ||   || ||
Db    1767021 TACAAATGTCAGACAAATCTCGTTTAATGAGTGANGATTATGGCATTGAGAAAATTATTG 1766962

Qy       1069 AGTGAGTGGGAGCGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGAT 1122
              |   |  |  || | || |||     ||| |   |   ||| || ||| ||| |
Db    1766961 AAGAATGGAAAGGGATTTTATAAGGNATAATTTTATGTTAAAAAAATATTTGNT 1766908


RESULT 7
US-09-489-039A-933
; Sequence 933, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342

; SEQ ID NO 933
;   LENGTH: 1095
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-933

  Query Match             5.1%;  Score 57.2;  DB 3;  Length 1095;
  Best Local Similarity   48.3%;  Pred. No. 0.00061;
  Matches  195;  Conservative    0;  Mismatches  203;  Indels    6;  Gaps    1;

Qy        655 TCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTA 714
              | |||   | |  | | |  | |   | | |  ||  | ||| |       |  ||| ||
Db        688 TTAAAAAATGACAAATATCATCTCACTCTAGTAGGACAAGGTGAGGATGATATGAAACTA 747

Qy        715 TTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGT 774
              ||  |  |   |  | |  || || |   |||||    ||  | |   ||| | |
Db        748 AGAGACATTTGTAAAAGACTAGATATTATAACAAAAGTCACTTTTATTGATAAAACAGAT 807

Qy        775 GATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGT 834
              |    | | |||||||      || ||| | ||| |||  || ||     ||||||||
Db        808 AACCCATATGCATTAATGAAACAAGCCGATGTATTTATTTCAAGCTCAAGATGGGAAGGA 867

Qy        835 TTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCT 894
              | |  |   ||||||   || ||| | || || | |   || |||||||| | |||| |
Db        868 TATCCTAATGTTGTTATTGAGGCTTTAGCATGCGGATTACCTGTTGTTGCCAATGATTAT 927

Qy        895 GGTGGAGTTA------AAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAAT 948
              | | ||      ||  | ||  ||    | ||     || |     | | || |||
Db        928 CCCGGGGGTATAAACGAAATAATCAATGATAGTAATGGCTATATATGTGATATAAGAAAT 987

Qy        949 CATATTCTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAA 1008
              || |              ||| ||  | | | | ||  |   | | |  |  |||
Db        988 GATCTAGAAGAAAATCTTGAAAGTGCACTAAAATTAAGAAATGTATCTTATGATAAGAAA 1047

Qy       1009 ATAATAGGTATGAAAAATAGAGAATATATTGTTTCCAATTTTTC 1052
                |  |||| | | |   |||| ||||   | | | | | ||||
Db       1048 AAGATAGATGTCATCTTTAGAGCATATGAAGATACTATTATTTC 1091


RESULT 8
US-08-961-527-155/c
; Sequence 155, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
```

```
;  INFORMATION FOR SEQ ID NO:  155:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6474 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-155

  Query Match             5.0%;  Score 56.8;  DB 3;  Length 6474;
  Best Local Similarity   46.9%;  Pred. No. 0.0013;
  Matches  212;  Conservative    0;  Mismatches  237;  Indels    3;  Gaps    1;

Qy        665 ATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGG 724
              | ||| | |||||  |||| |||||  ||||      |  |   ||    ||
Db       5892 AAGATGTGATTTTTATTATGGCTGGAAATGGTAGTATGTGGGCTGCTTTGAAAGAAAAAA 5833

Qy        725 TTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAG 784
              |||| || |||||    |  |||||  |    ||  | ||  |||    |     ||
Db       5832 TTTGCCATTTGAAATGTAGAGATAATTTCAGGTTACTTGGTGAAATCTATCCAGCTACTA 5773

Qy        785 AATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCG 844
              | |            | ||| | ||   | |   ||||||    |||||  |   | |
Db       5772 TAGTTTACCAAATATCGGATTTATTGTATATTCCTTCTGATACAGAAGGAATACCTATGT 5713

Qy        845 TTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAGTTA 904
              ||||  |||  | ||| |    | |  ||| ||| | ||||   ||  ||| ||  | |
Db       5712 GTGTTTTAGAGTCAATGTCACAAGGAACTCCGGTTTTAGCTAGTAATGTTGGCGGGTTAA 5653

Qy        905 AAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTGGCAG 964
              |||  |  ||| || |     |||||    |     |   ||  |   | |   || ||
Db       5652 GTGAAATTATTGAACATAGGGTTGATGGATTTTTATTTGAGAAGGAAGATGTTGAGGGAG 5593

Qy        965 AGAAAATCGCTGAGACACTTAAAATAGATGA---TAACGCAAGAAAAATAATAGGTATGA 1021
              |         ||  |   ||  | |  ||||   | |        |||   ||||||  ||
Db       5592 TGTGTGCTTGTGCTAATTTTTTACTCAATGATTCTGAGTATTTGAAATATATAGGTGAGA 5533

Qy       1022 AAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAGC 1081
              | | || |  | | ||        |||||||  |  |||  ||  |   ||       |
Db       5532 ATAGTAAATCAAAAATAAGAAAACATTTTTCTGTGCAAAAAATGTTTGTAGAAACCATGA 5473

Qy       1082 GCTTATATTTTAAATATTCCAAGCGTAATAAT 1113
              |  |||||  | |||  |   ||   | || |
Db       5472 GAGTATATGATGAATTATTAGAGAAGAGTAGT 5441


RESULT 9
US-09-583-110-2189
; Sequence 2189, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
;  APPLICANT:  Lynn Doucette-Stamm et al.
;  TITLE OF INVENTION:  Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
;  TITLE OF INVENTION:  Pneumoniae for Diagnostics and Therapeutics
;  FILE REFERENCE:  PATH00-07A
;  CURRENT APPLICATION NUMBER: US/09/583,110
;  CURRENT FILING DATE:  2000-05-26
;  PRIOR APPLICATION NUMBER:  US 09/107,433
;  PRIOR FILING DATE:  1998-06-30
;  PRIOR APPLICATION NUMBER:  US 60/085,131
;  PRIOR FILING DATE:  1998-05-12
;  PRIOR APPLICATION NUMBER:  US 60/051,553
;  PRIOR FILING DATE:  1997-07-02
;  NUMBER OF SEQ ID NOS:  5322
; SEQ ID NO 2189
;   LENGTH: 525
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-09-583-110-2189

  Query Match             4.9%;  Score 55.2;  DB 3;  Length 525;
  Best Local Similarity   46.7%;  Pred. No. 0.0015;
  Matches  211;  Conservative    0;  Mismatches  238;  Indels    3;  Gaps    1;

Qy        665 ATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGG 724
              | ||| | |||||  |||| |||||  |||      |  |   ||    ||
Db         65 AAGATGTGATTTTTATTATGGCTGGAAATGGTAGTATGTGGGCTGCTTTGAAAGAAAAAA 124

Qy        725 TTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAG 784
              |||| || |||||    |  |||||  |    ||  | ||  |||    |     ||
Db        125 TTTGCCATTTGAAATGTAGAGATAATTTCAGGTTACTTGGTGAAATCTATCCAGCTACTA 184

Qy        785 AATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCG 844
              | |            | ||| | ||   | |   || |||    |||||  |   | |
Db        185 TAGTTTACCAAATATCGGATTTATTGTATATTCCTTATGATACAGAAGGAATACCTATGT 244

Qy        845 TTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAGTTA 904
              ||||  |||  | ||| |    | |  ||| ||| | ||||   ||  ||| ||  | |
Db        245 GTGTTTTAGAGTCAATGTCACAAGGAACTCCGGTTTTAGCTAGTAATGTTGGCGGGTTAA 304

Qy        905 AAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTGGCAG 964
              |||  |  ||| || |     |||||    |     |   ||  |   | |   || ||
Db        305 GTGAAATTATTGAACATAGGGTTGATGGATTTTTATTTGAGAAGGAAGATGTTGAGGGAG 364

Qy        965 AGAAAATCGCTGAGACACTTAAAATAGATGA---TAACGCAAGAAAAATAATAGGTATGA 1021
              |         ||  |   ||  | |  ||||   | |        |||   ||||||  ||
Db        365 TGTGTGCTTGTGCTAATTTTTTACTCAATGATTCTGAGTATTTGAAATATATAGGTGAGA 424

Qy       1022 AAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAGC 1081
              | | || |  | | ||        |||||||  |  |||  ||  |   ||       |
Db        425 ATAGTAAATCAAAAATAAGAAAACATTTTTCTGTGCAAAAAATGTTTGTAGAAACCATGA 484

Qy       1082 GCTTATATTTTAAATATTCCAAGCGTAATAAT 1113
              |  |||||  | |||  |   ||   | || |
Db        485 GAGTATATGATGAATTATTAGAGAAGAGTAGT 516


RESULT 10
US-09-107-433-1123
; Sequence 1123, Application US/09107433
; Patent No. 6800744
;  GENERAL INFORMATION:
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                           SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGNO
;                           THERAPEUTICS
;       NUMBER OF SEQUENCES: 5206
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: <Unknown>
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: <Unknown>
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,433
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/ 085131
;            FILING DATE: May 12, 1998
;            APPLICATION NUMBER: 60/051553
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-011
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;  INFORMATION FOR SEQ ID NO: 1123:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 660 base pairs
```

```
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Streptococcus pneumoniae
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...660
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1123:
US-09-107-433-1123

  Query Match             4.9%;  Score 55.2;  DB 3;  Length 660;
  Best Local Similarity   46.7%;  Pred. No. 0.0016;
  Matches  211;  Conservative    0;  Mismatches  238;  Indels    3;  Gaps    1;

Qy        665 ATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGG 724
              | ||| | |||||  |||| |||||  ||||     |  |    ||   ||
Db        200 AAGATGTGATTTTTATTATGGCTGGAAATGGTAGTATGTGGGCTGCTTTGAAAGAAAAAA 259

Qy        725 TTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAG 784
              ||||  || |||||    |   |||||  |    ||  | ||  |||     |     ||
Db        260 TTTGCCATTTGAAATGTAGAGATAATTTCAGGTTACTTGGTGAAATCTATCCAGCTACTA 319

Qy        785 AATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCG 844
              | |              | ||| | ||   | |    || |||    |||||  |  | |
Db        320 TAGTTTACCAAATATCGGATTTATTGTATATTCCTTATGATACAGAAGGAATACCTATGT 379

Qy        845 TTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAGTTA 904
              ||||  |||  | ||| |    | |  ||| ||| | ||||   ||  ||| ||  | |
Db        380 GTGTTTTAGAGTCAATGTCACAAGGAACTCCGGTTTTAGCTAGTAATGTTGGCGGGTTAA 439

Qy        905 AAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTGGCAG 964
              |||  |  |||  ||  |    |||||    |    |    ||  |   |  |  || ||
Db        440 GTGAAATTATTGAACATAGGGTTGATGGATTTTTATTTGAGAAGGAAGATGTTGAGGGAG 499

Qy        965 AGAAAATCGCTGAGACACTTAAAATAGATGA---TAACGCAAGAAAAATAATAGGTATGA 1021
              |          ||  |   ||  | |  ||||    | |         |||   ||||||  ||
Db        500 TGTGTGCTTGTGCTAATTTTTTACTCAATGATTCTGAGTATTTGAAATATATAGGTGAGA 559

Qy       1022 AAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAGC 1081
              | | || |  | | ||         |||||||  |  |||  ||  |    ||      |
Db        560 ATAGTAAATCAAAAATAAGAAAACATTTTTCTGTGCAAAAAATGTTTGTAGAAACCATGA 619

Qy       1082 GCTTATATTTTAAATATTCCAAGCGTAATAAT 1113
              |  |||||  | |||  |    ||   | || |
Db        620 GAGTATATGATGAATTATTAGAGAAGAGTAGT 651


RESULT 11
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters

US-09-806-708B-22

  Query Match             4.9%;  Score 54.6;  DB 3;  Length 1141;
  Best Local Similarity   11.5%;  Pred. No. 0.0025;
  Matches  115;  Conservative  371;  Mismatches  500;  Indels   10;  Gaps    4;

Qy         63 GTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGTAT 122
              :::|: ::::  | : | | :: : | :::: :::   :|:::::  |    |: |   ::
Db         97 RYHARRWKDMKTAYBMTMTNKWGKTGWRHRYWRWRAMBDTVDHHYVTAMNNAWTTMCMMD 156

Qy        123 ATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAATTT 182
               : :: : ::      |  : ::  :: :      : :|:: ::: : :::  :   ::::
Db        157 KDDKRTRWWWKKNNNATGWDDDTKYHMWNNNGCBTVTWMVRYKTDRDWSBKRMNYGMBWW 216

Qy        183 TAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCAGA 242
              :    ::|:::: ::: |: ::  ::: : :::  ::|:  ::      :  |  ::::
Db        217 KNWSYDVTYYWWVWDDMCKRKVRRWVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMAYRRW 276

Qy        243 TATAATACATA-GTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTGA 301
              |:      |:|  : |: :|:      :|  |:: :|| :  : :      ||:| : :::
Db        277 TMNNNNNNAKAMCKRAKYWGWNRABVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDHKW 336

Qy        302 TTCCAGCGGTGCCCCTGATATGTACCGCACACAAC------AAAAATGAAGGTGGCAATG 355
              :: :| :|::         : :::| :: : : :::       :| ||:  :::: :: ::
Db        337 WKWSAAMGVYWNNNNNNNWTYKKARHBARWDWVWHSAWKKWHANAAHYSRKKWTBYKRKT 396

Qy        356 CAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTA--CTACAAATGTAAG 413
              ::       | ::: :|::  :: : :| :       | |:: :|:   :: :  ::    :
Db        397 MVNNNNGTTMWKRMWAWYWKMDMDWBGTYNNNNNGGRTYYGWTKNKKMWTYYKWKANNCK 456

Qy        414 TAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGA 473
              ::|:::: ||      :   :|::   :: ::  |:::::|      : :::::::|
Db        457 WRAWDHKTCTHNNTTWWKMKTYWNNCYWKSMTNGKSHRBAAAVYTWYMWWWRRYAHANNN 516

Qy        474 GATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAAC 533
               :::  :| |:::  :   ::      |: |: :: :| : ::::   :: :: : : :
Db        517 NWDYWWKACTWYKYBVCSKWWNNYAAWYTKSSWNYTSRYYRWKTNNSWRWRSDTRSMGRA 576

Qy        534 GCGAGATGCTTTTAATTTGAAAGACAGTAC-AGCAGTACTGCTCGCAGTAGGAAGACTTG 592
              |:|: : :::   :::: : :::: |  | :: ::::: :| |:: : :| :    |
Db        577 NNYARABHYGYKWNTRWWBWSHTWBHBRAGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAG 636

Qy        593 TTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAA 652
                                    | :: : | ::  :| | ::: ::     ::|:  ::
Db        637 GSNNNNNNNNNNNNNNNNATCARDDYYAASRWYAMANAKWYYYKBAANNAYYTHANNWWG 696

Qy        653 CATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAAT 712
              |:  | :| :|:::      | |::       :  :  :  :       :::   :
Db        697 CWNNATDTRRTMWKNNNNNNAGTWKNNNNNNAKNASAAKNYAAAAVKAAKKHWRWANKWA 756

Qy        713 TATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAA 772
               : :  :  : |:::   |:|:  |       : ::  |     :
Db        757 MRGWHADAAABTTDKRNNGAYTKYTTTNNNNTYRGVVTNTAARDGWANNNNNNNNNNNNN 816

Qy        773 GTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAG 832
               :::  ::  | :                | |: | ::: ||:::::: :|       ::|
Db        817 NGWSDMWVTWWAYANYGTNNNNNNNNNNAYAWWTNKWYYTTDDRWRBAYTNNNNNNRMAY 876

Qy        833 GTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATT 892
               : ::: ::  :  :  :::  ::      :  | | :|       | : : :: | :
Db        877 YGAYADDYAYYMSDTCDAWMKWDATKMNNATTYNRGTAWRTNNNNNNMTMKTKYYBHAAW 936

Qy        893 CTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATA 952
                  |  ||:  :: :  :| : |     |   : ::   | : : | :| :: ||:
Db        937 NNNNNNGKMCTAHTWWVCKATKTTKGCWMNCTTTCRKYKNNCTWYTWMTTTRTTWYAATR 996

Qy        953 TTCTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAA 1012
              ::    ||:: :  |  ::   :  |:   :|| ::  :: :|:: :    ::    :|:
Db        997 WKTNNATGSMTRCNATGWKNNNYWTGWKTRWTAYRMATRWMKAWWKVMATGSWNTNSYAR 1056

Qy       1013 TAGGTATGAAAAATAGAGAATATATTGTTTCCAATT 1048
              :| :|:    :   | | :: :   | : |  |::|
Db       1057 WAYKTRAYKGWYYNACAWRWRWGKATCYMTDNAWWT 1092
```

```
RESULT 12
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349491)..(1349491)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

  Query Match             4.7%;  Score 53.4;  DB 3;  Length 1664976;
  Best Local Similarity   53.0%;  Pred. No. 0.039;
  Matches  114;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        680 TTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATC 739
              | |||  ||| ||||| |  |||| ||||||| |    || ||  |   | |||| ||
Db     472412 TAATTCTTGGAGATGGGGAGTTAAAAAATAAACTTCAAGAATTAATAAATAAATTAAACT 472353

Qy        740 TTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAATTAATGTGTGCTG 799
              |    |||| ||||   | || ||      | ||      ||||  |      |
Db     472352 TACAAAATAATGTTTATCTTTTAGGCATGCAGAAGAATCCATTCAAATTTTAAAGCATT 472293

Qy        800 CAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCGTTGTTGCAGAAGCTA 859
              |  ||  |||||||||    |||  |||||| || ||      |  ||||  ||| ||
Db     472292 CGAATTGTTTTGTTTTCTCATCTTTGTGGGAGGGATTACCAAACACTGTTATAGAGGCGT 472233

Qy        860 TGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCT 894
              | |||   | |  || ||  |  | || |||| |
Db     472232 TATCGTTAAACCTCCCTGTAATATCAACTGATTGT 472198


RESULT 13
US-09-692-570-1/c
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6797466
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275C1
;  CURRENT APPLICATION NUMBER: US/09/692,570
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  PRIOR APPLICATION NUMBER: US 08/916,421
;  PRIOR FILING DATE: 1997-08-22
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g

Query Match 4.7%; Score 53.4; DB 3; Length 1664976;
Best Local Similarity 53.0%; Pred. No. 0.039;
Matches 114; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```
Qy        680 TTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATC 739
              | |||  ||| ||||| |  |||| |||||||  |    || ||  |   | |||| ||
Db     472412 TAATTCTTGGAGATGGGGAGTTAAAAAATAAACTTCAAGAATTAATAAATAAATTAAACT 472353

Qy        740 TTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAATTAATGTGTGCTG 799
              |    |||| ||||   | || ||       |   ||         ||||  |      |
Db     472352 TACAAAATAATGTTTATCTTTTAGGCATGCAGAAGAATCCATTCAAATTTTTAAAGCATT 472293

Qy        800 CAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCGTTGTTGCAGAAGCTA 859
              |  ||  |||||||||    |||  |||||| || ||      |  ||||  ||| ||
Db     472292 CGAATTGTTTTGTTTTCTCATCTTTGTGGGAGGGATTACCAAACACTGTTATAGAGGCGT 472233

Qy        860 TGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCT 894
              |  |||   | |  || ||  |  | || |||| |
Db     472232 TATCGTTAAACCTCCCTGTAATATCAACTGATTGT 472198
```

RESULT 14
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22

; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)

```
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

  Query Match             4.7%;  Score 52.8;  DB 3;  Length 1664976;
  Best Local Similarity   46.9%;  Pred. No. 0.054;
  Matches  199;  Conservative    0;  Mismatches  222;  Indels    3;  Gaps    1;

Qy        640 ATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCA 699
              ||| |   | |  ||  |||   ||  ||||||| |||     |    || |||||
Db     355489 ATTTTAATAGATGCAATAAAAGATATAGATTTTAATTTTAAACTCATAGGGGATGGGAAG 355548

Qy        700 TTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTC 759
              |||   || ||| ||  | | || |||  | ||   ||| |     |||  |   ||||
Db     355549 TTATACAAAAAAATAGAGAACTTTGTTGTTAAAAATAATTTAAGCCATATTGAACTCTTA 355608

Qy        760 TTGGGGCAAAGAAGTGATATTAAAG---AATTAATGTGTGCTGCAGATCTTTTTGTTTTG 816
                | ||||   ||||     ||    ||| ||| |        | |||| ||  |
Db     355609 GGAAGAAAAAGTTTTGATGAAGTAGCTTCATTTATGAGGAAGTGTAGTTTTTTAGTAGTT 355668

Qy        817 AGTTCTGAGTGGGAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCC 876
              |||     | ||||||||||  | || | ||  ||||  ||||| ||       ||
Db     355669 CCTTCAAGAAGTGAAGGTTTTGGAATGGTGGCTGTTGAAGGAATGGCTTGCTCTAAGCCT 355728

Qy        877 GTTGTTGCTACCGATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATC 936
              ||  | || ||       ||| ||  |    ||  | |||     |    || | |   |
Db     355729 GTAATAGCCACAAGGGTTGGGGGGTTGGGGGAGATTGTTATTGATGGATATAACGGACTA 355788

Qy        937 CCTGTCAGTAATCATATTCTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGAT 996
              |     ||  |||  |     |    |||                  |||| || ||
Db     355789 TTGGCTGAGAAAAATAACCCAAATGATTTAAAAGAAAAAATTCTGGAGTTAATAAATAAT 355848

Qy        997 AACGCAAGAAAAATAATAGGTATGAAAAATAGAGAATATATTGTTTCCAATTTTTCAATT 1056
              | | |  || || ||       | ||||| || || |  ||      || |||||
Db     355849 GAAGAACTAAGAAAAACTTTGGGGGAAAATGGAAAAGAATTTTCAAAAAAATTTTCTTGG 355908

Qy       1057 AAAA 1060
                |||
Db     355909 GAAA 355912


RESULT 15
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6797466
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275C1
;  CURRENT APPLICATION NUMBER: US/09/692,570
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  PRIOR APPLICATION NUMBER: US 08/916,421
;  PRIOR FILING DATE: 1997-08-22
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
```

; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g

Query Match 4.7%; Score 52.8; DB 3; Length 1664976;
Best Local Similarity 46.9%; Pred. No. 0.054;
Matches 199; Conservative 0; Mismatches 222; Indels 3; Gaps 1;

```
Qy       640 ATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCA 699
Db    355489 ATTTTAATAGATGCAATAAAAGATATAGATTTTAATTTTAAACTCATAGGGGATGGGAAG 355548

Qy       700 TTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTC 759
Db    355549 TTATACAAAAAAATAGAGAACTTTGTTGTTAAAAATAATTTAAGCCATATTGAACTCTTA 355608

Qy       760 TTGGGGCAAAGAAGTGATATTAAAG---AATTAATGTGTGCTGCAGATCTTTTTGTTTTG 816
Db    355609 GGAAGAAAAAGTTTTGATGAAGTAGCTTCATTTATGAGGAAGTGTAGTTTTTTAGTAGTT 355668

Qy       817 AGTTCTGAGTGGGAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCC 876
Db    355669 CCTTCAAGAAGTGAAGGTTTTGGAATGGTGGCTGTTGAAGGAATGGCTTGCTCTAAGCCT 355728

Qy       877 GTTGTTGCTACCGATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATC 936
Db    355729 GTAATAGCCACAAGGGTTGGGGGGTTGGGGGAGATTGTTATTGATGGATATAACGGACTA 355788

Qy       937 CCTGTCAGTAATCATATTCTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGAT 996
Db    355789 TTGGCTGAGAAAAATAACCCAAATGATTTAAAAGAAAAAATTCTGGAGTTAATAAATAAT 355848

Qy       997 AACGCAAGAAAAATAATAGGTATGAAAAATAGAGAATATATTGTTTCCAATTTTTCAATT 1056
Db    355849 GAAGAACTAAGAAAAACTTTGGGGGAAAATGGAAAAGAATTTTCAAAAAAATTTTCTTGG 355908
```

```
Qy       1057 AAAA 1060
                 |||
Db     355909 GAAA 355912
```

Search completed: May 11, 2006, 05:23:42
Job time : 171.922 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:01:11 ; Search time 742.533 Seconds
(without alignments)
12528.801 Million cell updates/sec

Title: US-09-423-093B-1_COPY_11821_12945
Perfect score: 1125
Sequence: 1 atgaagattgtttatataat.....gtaataatataattgattga 1125

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1125	100.0	14516	3	US-09-423-093A-1	Sequence 1, Appli
	2	245.8	21.8	1104	6	US-10-238-075-1321	Sequence 1321, Ap
	3	245.8	21.8	4604	10	US-11-040-661-22	Sequence 22, Appl
c	4	245.8	21.8	6927	6	US-10-238-075-1318	Sequence 1318, Ap
	5	73.8	6.6	1143	7	US-10-282-122A-10729	Sequence 10729, A
	6	73.8	6.6	1156	7	US-10-282-122A-9697	Sequence 9697, Ap
	7	65.6	5.8	1116	8	US-10-472-928-537	Sequence 537, App
	8	65.6	5.8	1128	7	US-10-474-776-19	Sequence 19, Appl
c	9	65.6	5.8	19250	2	US-08-961-527-35	Sequence 35, Appl
c	10	65.6	5.8	19250	7	US-10-158-844-35	Sequence 35, Appl
	11	65.6	5.8	2162598	8	US-10-472-928-4979	Sequence 4979, Ap
c	12	63.6	5.7	3262	9	US-10-795-159-409	Sequence 409, App
c	13	63.6	5.7	908766	9	US-10-795-159-685	Sequence 685, App
	14	62.6	5.6	1092	7	US-10-282-122A-16045	Sequence 16045, A
	15	60.6	5.4	1101	6	US-10-369-493-34143	Sequence 34143, A
	16	60	5.3	3089	9	US-10-954-778-160	Sequence 160, App
	17	60	5.3	4985	5	US-10-094-240-10	Sequence 10, Appl
	18	60	5.3	4985	6	US-10-056-405-10	Sequence 10, Appl
c	19	58	5.2	1830121	7	US-10-329-670-1	Sequence 1, Appli
c	20	58	5.2	1830121	8	US-10-158-865-1	Sequence 1, Appli
c	21	58	5.2	1830121	9	US-10-981-687-1	Sequence 1, Appli
c	22	56.8	5.0	6474	2	US-08-961-527-155	Sequence 155, App
c	23	56.8	5.0	6474	7	US-10-158-844-155	Sequence 155, App
c	24	55.4	4.9	351	9	US-10-501-282-3539	Sequence 3539, Ap
	25	55.4	4.9	1185	9	US-10-501-282-3535	Sequence 3535, Ap
	26	55.4	4.9	1185	9	US-10-501-282-3537	Sequence 3537, Ap
c	27	55.4	4.9	1754382	9	US-10-501-282-6651	Sequence 6651, Ap
	28	55.2	4.9	660	9	US-10-617-320-1123	Sequence 1123, Ap
	29	54	4.8	453	8	US-10-472-928-67	Sequence 67, Appl
	30	53.8	4.8	3673778	6	US-10-312-841-2	Sequence 2, Appli
	31	52.2	4.6	4318	10	US-11-097-143-10244	Sequence 10244, A
	32	52.2	4.6	6318	10	US-11-097-143-10243	Sequence 10243, A
c	33	52.2	4.6	8056	8	US-10-473-126-386	Sequence 386, App
	34	52	4.6	6032	7	US-10-240-454-6	Sequence 6, Appli
	35	52	4.6	3673778	6	US-10-312-841-1	Sequence 1, Appli
c	36	51.6	4.6	3089	9	US-10-954-778-160	Sequence 160, App
c	37	51.6	4.6	4985	5	US-10-094-240-10	Sequence 10, Appl
c	38	51.6	4.6	4985	6	US-10-056-405-10	Sequence 10, Appl
c	39	51.4	4.6	588	7	US-10-437-963-73933	Sequence 73933, A
	40	51.4	4.6	1173	6	US-10-369-493-24780	Sequence 24780, A
	41	51.4	4.6	2020	4	US-09-925-065A-728360	Sequence 728360,
	42	51.2	4.6	1143	7	US-10-282-122A-8119	Sequence 8119, Ap
	43	51.2	4.6	1143	8	US-10-857-625-322	Sequence 322, App
	44	51.2	4.6	6025	8	US-10-857-625-76	Sequence 76, Appl
	45	50.8	4.5	9326	6	US-10-138-701-51	Sequence 51, Appl

ALIGNMENTS

RESULT 1
US-09-423-093A-1
; Sequence 1, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 14516
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-423-093A-1

Query Match 100.0%; Score 1125; DB 3; Length 14516;
Best Local Similarity 100.0%; Pred. No. 8.9e-218;
Matches 1125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11821 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 11880

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11881 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 11940

Qy        121 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11941 ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 12000

Qy        181 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12001 TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCA 12060

Qy        241 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12061 GATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTG 12120
```

```
Qy      301 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 360
Db    12121 ATTCCAGCGGTGCCCCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGG 12180

Qy      361 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 420
Db    12181 ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 12240

Qy      421 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 480
Db    12241 GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 12300

Qy      481 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540
Db    12301 AATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 12360

Qy      541 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 600
Db    12361 GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 12420

Qy      601 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 660
Db    12421 AAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 12480

Qy      661 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 720
Db    12481 TGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 12540

Qy      721 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 780
Db    12541 TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATT 12600

Qy      781 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 840
Db    12601 AAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGT 12660

Qy      841 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 900
Db    12661 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGA 12720

Qy      901 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 960
Db    12721 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG 12780

Qy      961 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 1020
Db    12781 GCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG 12840

Qy     1021 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 1080
Db    12841 AAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAG 12900

Qy     1081 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 1125
Db    12901 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGATTGA 12945
```

```
RESULT 2
US-10-238-075-1321
; Sequence 1321, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
;  APPLICANT: I.N.S.E.R.M.
;  TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat
;  TITLE OF INVENTION:  E.coli, and biological uses of these polynucleotides and of thei
;  FILE REFERENCE: BLANDINE
;  CURRENT APPLICATION NUMBER: US/10/238,075
;  CURRENT FILING DATE:  2002-09-10
;  PRIOR APPLICATION NUMBER: 0003145
;  PRIOR FILING DATE: 2000-03-10
;  NUMBER OF SEQ ID NOS: 1576
;  SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 1321
;   LENGTH: 1104
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-10-238-075-1321

  Query Match             21.8%;  Score 245.8;  DB 6;  Length 1104;
  Best Local Similarity   55.8%;  Pred. No. 9e-40;
  Matches  550;  Conservative    0;  Mismatches  402;  Indels   33;  Gaps    3;

Qy        1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
Db        1 ATGAAAATACTATTTGTCATTACAGGTTTAGGCCTTGGAGGTGCTGAGAAGCAGGTTTGT 60

Qy       61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
Db       61 CTTTTAGCTGATAAATTAAGTTTAAGCGGGCACCATGTAAAGATTATTTCACTTGGACAT 120

Qy      121 ATATCTGAGGTAAAGCCAACAC---------AAAATATTAATATTCATTATGTTAATATG 171
Db      121 ATGTCTAATAATAAAGTCTTTCCTAGCGAAAATAATGTTAATGTCATTAATGTAAATATG 180

Qy      172 GATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCC 231
Db      181 TCAAAAAACATTTCTGGAGTTATAAAAGGTTGTGTCAGAATTAGAGATGTTATAGCTAAT 240

Qy      232 TTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATT 291
Db      241 TTCAAACCAGACATTGTACACAGTCATATGTTTCATGCAAACATTATCACTAGATTGTCT 300

Qy      292 AGGATGCTGATTCCAGCGGTGCC---CCTGATATGTACCGCACACAACAAAAATGAAGGT 348
Db      301 GTAATTGGAATCAAAAACAGACCTGGTATTATATCAACTGCACATAATAAAAATGAAGGT 360

Qy      349 GGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAAT 408
Db      361 GGGTATTTCAGAATGCTCACATATAGAATAACCGATTGTTTAAGTGATTGTTGTACAAAT 420

Qy      409 GTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATA 468
Db      421 GTTAGCAAAGAAGCAGTGGATGAGTTTTTACGGATAAAAGCCTTTAATCCCGCTAAAGCA 480

Qy      469 GTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAG 528
Db      481 ATTACTATGTATAATGGGATAGATACCAATAAATTTAAATTTGATTTATTGGCAAGGAGG 540

Qy      529 AAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGA 588
Db      541 GAAATTCGAGACGGTATTAATATAAAAAATGATGATATATTATTACTTGCTGCAGGTCGT 600

Qy      589 CTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCA 648
Db      601 TTAACGTTAGCTAAAGATTATCCTAATTTATTGAATGCAATGACTCTGCT---------- 650

Qy      649 AAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAAT 708
Db      651 ----------TCCTGAACACTTTAAACTTATTATTATTGGTGATGGTGAATTGCGTGAC 699

Qy      709 AAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAA 768
Db      700 GAAATTAATATGCTTATAAAAAAATTGCAATTATCTAATAGGGTGTCCTTGTTGGGAGTT 759

Qy      769 AGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGG 828
Db      760 AAAAAAAATATTGCTCCCTATTTTTCTGCATGTGATATTTTTGTTCTCTCTTCTCGTTGG 819

Qy      829 GAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACC 888
Db      820 GAAGGATTTGGATTAGTCGTGGCAGAAGCTATGTCATGTGAGCGAATTGTTGTTGGCACG 879

Qy      889 GATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAAT 948
Db      880 GATTCAGGGGGAGTAAGAGAAGTTATTGGTGACGATGATTTTCTTGTACCCATATCTGAT 939
```

```
Qy        949 CATATTCTGTTGGCAGAGAAAATCG 973
              |  |   | |||   ||||| |
Db        940 TCAACACAACTTGCAAGCAAAATTG 964
```

RESULT 3
US-11-040-661-22
; Sequence 22, Application US/11040661
; Publication No. US20050142149A1
; GENERAL INFORMATION:
; APPLICANT: Crooke,Helen R.
; APPLICANT: Clarke, Enda E.
; APPLICANT: Everest, Paul H.
; APPLICANT: Dougan, Gordon
; APPLICANT: Holden, David W.
; APPLICANT: Shea, Jacqueline E.
; APPLICANT: Feldman, Robert G.
; TITLE OF INVENTION: VIRULENCE GENES AND PROTEINS, AND THEIR USE
; FILE REFERENCE: GJE-65
; CURRENT APPLICATION NUMBER: US/11/040,661
; CURRENT FILING DATE: 2005-01-21
; PRIOR APPLICATION NUMBER: US/09/830,807
; PRIOR FILING DATE: 2001-07-16
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 4604
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2982)..(4082)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1534)..(2637)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (749)..(1531)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (6)..(746)
US-11-040-661-22

Query Match 21.8%; Score 245.8; DB 10; Length 4604;
Best Local Similarity 55.8%; Pred. No. 1.6e-39;
Matches 550; Conservative 0; Mismatches 402; Indels 33; Gaps 3;

```
Qy          1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||| ||  | | | | || || ||  |     ||| || || ||  | |  |
Db       1534 ATGAAAATACTATTTGTCATTACAGGTTTAGGCCTTGGAGGTGCTGAGAAGCAGGTTTGT 1593

Qy         61 CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 120
              |  ||||| ||  || |   | || |||||||  |||| || |||||||  ||       |
Db       1594 CTTTTAGCTGATAAATTAAGTTTAAGCGGGCACCATGTAAAGATTATTTCACTTGGACAT 1653

Qy        121 ATATCTGAGGTAAAGCCAACAC---------AAAATATTAATATTCATTATGTTAATATG 171
              || ||| |    ||       |         | ||| ||||| |   | |||| ||||||
Db       1654 ATGTCTAATAATAAAGTCTTTCCTAGCGAAAATAATGTTAATGTCATTAATGTAAATATG 1713

Qy        172 GATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCC 231
              |||||  ||    |  || | | || ||   |  | | | | |  | || |
Db       1714 TCAAAAAACATTTCTGGAGTTATAAAAGGTTGTGTCAGAATTAGAGATGTTATAGCTAAT 1773

Qy        232 TTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTTTTAGTCGTTTTATT 291
              || || ||||| ||  |||| ||||||||||||||||| || ||| | | | ||   |
Db       1774 TTCAAACCAGACATTGTACACAGTCATATGTTTCATGCAAACATTATCACTAGATTGTCT 1833

Qy        292 AGGATGCTGATTCCAGCGGTGCC---CCTGATATGTACCGCACACAACAAAAATGAAGGT 348
              ||    ||   |      ||     | ||||  || ||||| || ||||||||||||
Db       1834 GTAATTGGAATCAAAAACAGACCTGGTATTATATCAACTGCACATAATAAAAATGAAGGT 1893

Qy        349 GGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAAT 408
```

```
              || ||   || ||| |    ||| || | |  |||| ||||  |  |  |  |||||||
Db       1894 GGGTATTTCAGAATGCTCACATATAGAATAACCGATTGTTTAAGTGATTGTTGTACAAAT 1953

Qy        409 GTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATA 468
              || || ||||| || ||  | |||||| ||   | ||| ||       |     ||||  |
Db       1954 GTTAGCAAAGAAGCAGTGGATGAGTTTTTACGGATAAAAGCCTTTAATCCCGCTAAAGCA 2013

Qy        469 GTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAG 528
              |    ||    |||   ||  |||| |||||||||  | ||||||  |    |  || | |
Db       2014 ATTACTATGTATAATGGGATAGATACCAATAAATTTAAATTTGATTTATTGGCAAGGAGG 2073

Qy        529 AAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGA 588
              |||  ||||| | | ||||| | ||| |   |      | || | || || | |||  |
Db       2074 GAAATTCGAGACGGTATTAATATAAAAAATGATGATATATTATTACTTGCTGCAGGTCGT 2133

Qy        589 CTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCA 648
              |      ||| ||||| ||||| || ||||| |||||||| | ||   |
Db       2134 TTAACGTTAGCTAAAGATTATCCTAATTTATTGAATGCAATGACTCTGCT---------- 2183

Qy        649 AAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAAT 708
                      | | | | ||||   |  |||||  ||| ||||| | |||  |  |
Db       2184 ----------TCCTGAACACTTTAAACTTATTATTATTGGTGATGGTGAATTGCGTGAC 2232

Qy        709 AAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAA 768
              || |        |  |     ||||| |  |     |||  || | ||| |||||
Db       2233 GAAATTAATATGCTTATAAAAAAAATTGCAATTATCTAATAGGGTGTCCTTGTTGGGAGTT 2292

Qy        769 AGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGG 828
              | ||   |||||       |   | | |||     ||| |||||||| |   ||||   |||
Db       2293 AAAAAAAATATTGCTCCCTATTTTTCTGCATGTGATATTTTTGTTCTCTCTTCTCGTTGG 2352

Qy        829 GAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACC 888
              ||||| |||||  | || || ||||||||||||| | |||||  ||    |||||||  ||
Db       2353 GAAGGATTTGGATTAGTCGTGGCAGAAGCTATGTCATGTGAGCGAATTGTTGTTGGCACG 2412

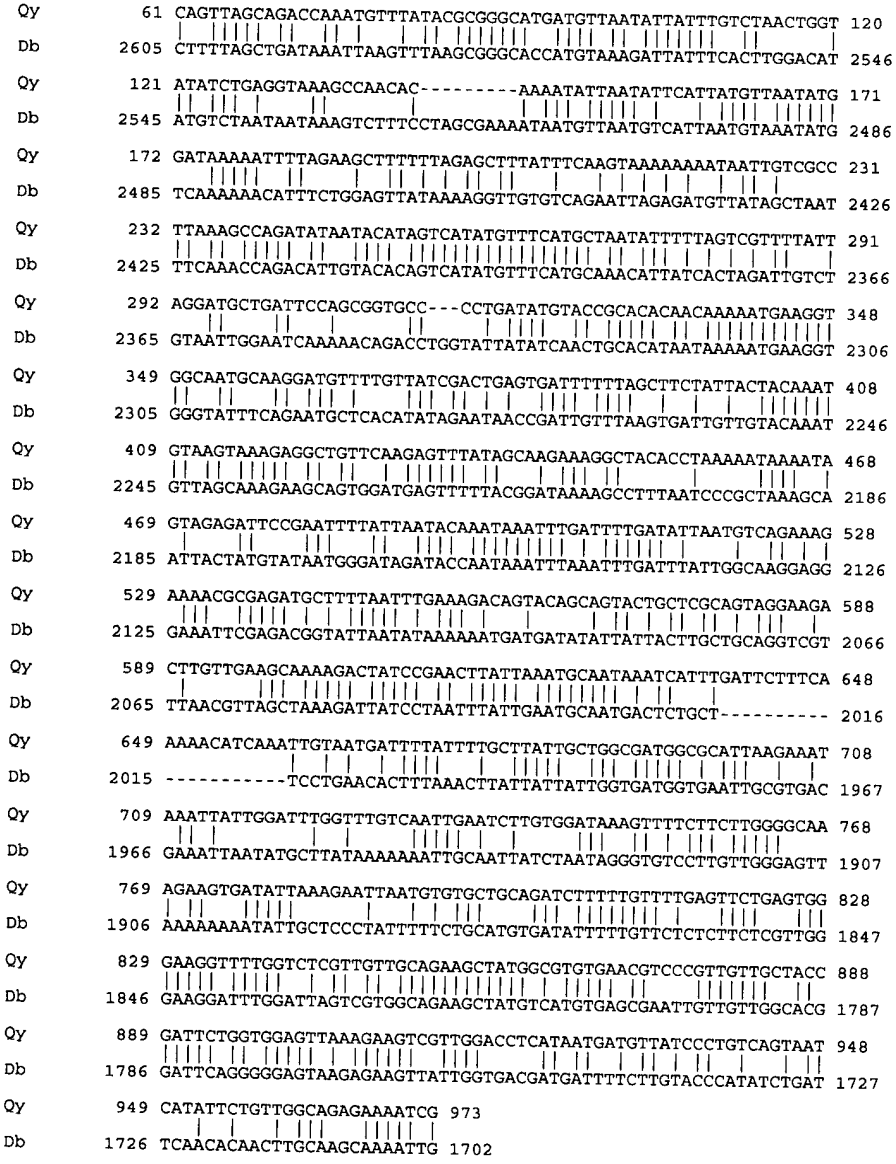
Qy        889 GATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAAT 948
              ||||| || ||||| | ||||||  ||||      || ||  | || | ||  |    | ||
Db       2413 GATTCAGGGGGAGTAAGAGAAGTTATTGGTGACGATGATTTTCTTGTACCCATATCTGAT 2472

Qy        949 CATATTCTGTTGGCAGAGAAAATCG 973
              |  |   | |||   ||||| |
Db       2473 TCAACACAACTTGCAAGCAAAATTG 2497

RESULT 4
US-10-238-075-1318/c
; Sequence 1318, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of thei
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1318
; LENGTH: 6927
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-1318

Query Match 21.8%; Score 245.8; DB 6; Length 6927;
Best Local Similarity 55.8%; Pred. No. 2e-39;
Matches 550; Conservative 0; Mismatches 402; Indels 33; Gaps 3;

```
Qy   1 ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 60
              ||||| || | | | | || || || | ||| || || || | | |
Db   2665 ATGAAAATACTATTTGTCATTACAGGTTTAGGCCTTGGAGGTGCTGAGAAGCAGGTTTGT 2606
```
```



RESULT 5
US-10-282-122A-10729
; Sequence 10729, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10729
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Bacillus anthracis
US-10-282-122A-10729

Query Match 6.6%; Score 73.8; DB 7; Length 1143;
Best Local Similarity 55.2%; Pred. No. 7.6e-05;
Matches 144; Conservative 0; Mismatches 117; Indels 0; Gaps 0;



RESULT 6
US-10-282-122A-9697
; Sequence 9697, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu

```
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9697
;   LENGTH: 1156
;   TYPE: DNA
;   ORGANISM: Bacillus anthracis
US-10-282-122A-9697

  Query Match             6.6%;  Score 73.8;  DB 7;  Length 1156;
  Best Local Similarity   55.2%;  Pred. No. 7.6e-05;
  Matches  144;  Conservative    0;  Mismatches  117;  Indels    0;  Gaps    0;

Qy        667 GATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTT 726
              |||   |  |||||| ||| ||| |||||  ||  |       |   | ||  | || ||
Db        703 GATGCGAAGTTGCTTCTTGTTGGAGATGGACCAGAATTTTGCACGATTTTACAGTTAGTG 762

Qy        727 TGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAA 786
              | ||  || ||| |||   || || ||||| ||| |   |  | || ||  ||||
Db        763 AAAAATTTACATATTGAAGATCGCGTCTTATTCTTAGGGAAGCAAGATAATGTTGCAGAA 822

Qy        787 TTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCGTT 846
               || |          ||| |  |  |  |    || |||  |||  ||||||||||| |||
Db        823 CTACTTGCGATGAGTGATTTAATGTTGCTTTTATCAGAGAAGGAGAGTTTTGGTCTTGTT 882

Qy        847 GTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAGTTAAA 906
              |   |||||| ||||||||||     ||   | | |  ||      ||| ||  ||  |
Db        883 TTATTAGAAGCGATGGCGTGTGGTGTACCTTGTATCGGAACAAGGGTTGGAGGTATTCCA 942

Qy        907 GAAGTCGTTGGACCTCATAAT 927
              |||||| ||  || |  | ||
Db        943 GAAGTCATTCAACATGGTGAT 963


RESULT 7
US-10-472-928-537
; Sequence 537, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
;  TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE: P026926WO
;  CURRENT APPLICATION NUMBER: US/10/472,928
;  CURRENT FILING DATE:  2003-09-26
;  PRIOR APPLICATION NUMBER: GB-0107658.7
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 4979
;  SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 537
;   LENGTH: 1116
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-10-472-928-537

  Query Match             5.8%;  Score 65.6;  DB 8;  Length 1116;
  Best Local Similarity   47.8%;  Pred. No. 0.0035;
  Matches  265;  Conservative    0;  Mismatches  274;  Indels   15;  Gaps    2;

Qy        416 AAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGA 475
              ||||  | | | |   |||    | |   || ||         |    ||||| | |   |
Db        437 AAGATTCGGCTAAGTGGTTATATGGAGAGAAAGCGTATAAAGACGGAAAAATTGAAATTA 496

Qy        476 TTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGC 535
              ||   |||   ||| ||   || || | | |||       |||| ||| | |||   |
Db        497 TTTTTAATGCAATTGATTTGAAAAAGTATCAATTTTTGTCAGATGTTAGAGAAAAATGTC 556

Qy        536 GAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTG 595
              |    |  ||  || ||  |   | ||     ||| |         |||   | ||    |
Db        557 GTAGAGAATTAGATGTGTCA---AATAAGTTCGTATTAGGAAATATAGCTCGCCTATCAG 613

Qy        596 AAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACAT 655
              |   |||| || |     | ||||| || |   ||||  |  | ||| |    ||| ||
Db        614 ATCAAAAAAACCAAAGTTATTTATTAACGTTTTAAAAGAACTCATTTTAATCAAACCA- 672

Qy        656 CAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTAT 715
                      || ||||||| ||  | | |||  |||| |          | ||||||
Db        673 ----------AATGTTATTTTACTCCTAGTTGGTAATGGTGAGGATGAGCAGAAATTAA 721

Qy        716 TGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTG 775
                  | ||   || | |||||         | ||  | || || |||  ||| | ||
Db        722 AACAGAAAGCTTTAGAACTAAATCTGACCCCATATGTGCTATTTTTAGGGAGAAGGACTG 781

Qy        776 ATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTT 835
              |||||   || ||| | | |||     ||| ||||| |  |    ||| | |  || ||| |
Db        782 ATATTTCTGATTTATTATCTGCGATGGATGTTTTTTTGCTTCCGTCTAAATATGAGGGGT 841

Qy        836 TTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTG 895
              |   | |   | | | ||| |||  ||| |  | |    |   || |  | ||| | |
Db        842 TGCCTGTTTCTCTAGTAGAGGCTCAGGCATCGGGATTACAAATTTTATCGTCAGATACAG 901

Qy        896 GTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTC 955
                  ||| ||  ||        | | |  |    | ||| |||   |  |   |
Db        902 TGACGCAAGAAGTAGATGTGACCAAAAACATTAGTTACTTACCTATCAACGAAGAGTCTG 961

Qy        956 TGTTGGCAGAGAAA 969
              |||||  |  ||||
Db        962 TGTTGCTATGGAAA 975


RESULT 8
US-10-474-776-19
; Sequence 19, Application US/10474776
; Publication No. US20040110181A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  TITLE OF INVENTION: NOVEL STREPTOCOCCUS PNEUMONIAE OPEN READING FRAMES ENCODING POLYP
;  TITLE OF INVENTION:  ANTIGENS AND USES THEREOF
;  FILE REFERENCE: AM100649-PCT
;  CURRENT APPLICATION NUMBER: US/10/474,776
;  CURRENT FILING DATE:  2003-10-14
```

```
;  NUMBER OF SEQ ID NOS: 752
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
;   LENGTH: 1128
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-10-474-776-19

  Query Match             5.8%;  Score 65.6;  DB 7;  Length 1128;
  Best Local Similarity   47.8%;  Pred. No. 0.0035;
  Matches  265;  Conservative    0;  Mismatches  274;  Indels   15;  Gaps    2;

Qy        416 AAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGA 475
Db        446 AAGATTCGGCTAAGTGGTTATATGGAGAGAAAGCGTATAAAGACGGAAAAATTGAAATTA 505

Qy        476 TTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGC 535
Db        506 TTTTTAATGCAATTGATTTGAAAAAGTATCAATTTTTGTCAGATGTTAGAGAAAAATGTC 565

Qy        536 GAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTG 595
Db        566 GTAGAGAATTAGATGTGTCA---AATAAGTTCGTATTAGGAAATATAGCTCGCCTATCAG 622

Qy        596 AAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACAT 655
Db        623 ATCAAAAAAACCAAAGTTATTTATTTAACGTTTTAAAAGAACTCATTTTAATCAAACCA- 681

Qy        656 CAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTAT 715
Db        682 -----------AATGTTATTTTACTCCTAGTTGGTAATGGTGAGGATGAGCAGAAATTAA 730

Qy        716 TGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTG 775
Db        731 AACAGAAAGCTTTAGAACTAAATCTGACCCCATATGTGCTATTTTTAGGGAGAAGGACTG 790

Qy        776 ATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTT 835
Db        791 ATATTTCTGATTTATTATCTGCGATGGATGTTTTTTTGCTTCCGTCTAAATATGAGGGGT 850

Qy        836 TTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTG 895
Db        851 TGCCTGTTTCTCTAGTAGAGGCTCAGGCATCGGGATTACAAATTTTATCGTCAGATACAG 910

Qy        896 GTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTC 955
Db        911 TGACGCAAGAAGTAGATGTGACCAAAAACATTAGTTACTTACCTATCAACGAAGAGTCTG 970

Qy        956 TGTTGGCAGAGAAA 969
Db        971 TGTTGCTATGGAAA 984


RESULT 9
US-08-961-527-35/c
; Sequence 35, Application US/08961527
; Publication No. US20020032323A1
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19250 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-35

  Query Match             5.8%;  Score 65.6;  DB 2;  Length 19250;
  Best Local Similarity   47.8%;  Pred. No. 0.011;
  Matches  265;  Conservative    0;  Mismatches  274;  Indels   15;  Gaps    2;

Qy        416 AAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGA 475
Db      11753 AAGATTCGGCTAAGTGGTTATATGGAGAGAAAGCGTATAAAGACGGAAAAATTGAAATTA 11694

Qy        476 TTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGC 535
Db      11693 TTTTTAATGCAATTGATTTGAAAAAGTATCAATTTTTGTCAGATGTTAGAGAAAAATGTC 11634

Qy        536 GAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTG 595
Db      11633 GTAGAGAATTAGATGTGTCA---AATAAGTTCGTATTAGGAAATATAGCTCGCCTATCAG 11577

Qy        596 AAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACAT 655
Db      11576 ATCAAAAAAACCAAAGTTATTTATTTAACGTTTTAAAAGAACTCATTTTAATCAAACCA- 11518

Qy        656 CAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTAT 715
Db      11517 -----------AATGTTATTTTACTCCTAGTTGGTAATGGTGAGGATGAGCAGAAATTAA 11469

Qy        716 TGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTG 775
Db      11468 AACAGAAAGCTTTAGAACTAAATCTGACCCCATATGTGCTATTTTTAGGGAGAAGGACTG 11409

Qy        776 ATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTT 835
Db      11408 ATATTTCTGATTTATTATCTGCGATGGATGTTTTTTTGCTTCCGTCTAAATATGAGGGGT 11349

Qy        836 TTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTG 895
Db      11348 TGCCTGTTTCTCTAGTAGAGGCTCAGGCATCGGGATTACAAATTTTATCGTCAGATACAG 11289

Qy        896 GTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTC 955
Db      11288 TGACGCAAGAAGTAGATGTGACCAAAAACATTAGTTACTTACCTATCAACGAAGAGTCTG 11229

Qy        956 TGTTGGCAGAGAAA 969
Db      11228 TGTTGCTATGGAAA 11215


RESULT 10
US-10-158-844-35/c
; Sequence 35, Application US/10158844
; Publication No. US20040029118A1
;  GENERAL INFORMATION:
;       APPLICANT: Kunsch et al.
```

```
;       TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
;       NUMBER OF SEQUENCES: 391
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: Maryland
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD-R
;            COMPUTER: Dell Latitude Pentium 3
;            OPERATING SYSTEM: Windows 98
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/158,844
;            FILING DATE: 03-Jun-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 08/961,527
;            FILING DATE: 1997-10-30
;            APPLICATION NUMBER: US 60/029,960
;            FILING DATE: 1996-10-31
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Hyman, Mark J.
;            REGISTRATION NUMBER: 46,789
;            REFERENCE/DOCKET NUMBER: PB340P1D1
;  INFORMATION FOR SEQ ID NO: 35:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 19250 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-10-158-844-35

  Query Match             5.8%;  Score 65.6;  DB 7;  Length 19250;
  Best Local Similarity   47.8%;  Pred. No. 0.011;
  Matches  265;  Conservative    0;  Mismatches  274;  Indels   15;  Gaps    2;

Qy        416 AAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGA 475
              |||| | | | |   |||    | |   || ||         |    ||||| | |   |
Db      11753 AAGATTCGGCTAAGTGGTTATATGGAGAGAAAGCGTATAAAGACGGAAAAATTGAAATTA 11694

Qy        476 TTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGC 535
              ||   |||   ||| ||   || || | | | |||       |||| ||| | |||   |
Db      11693 TTTTTAATGCAATTGATTTGAAAAAGTATCAATTTTTGTCAGATGTTAGAGAAAAATGTC 11634

Qy        536 GAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTG 595
              |    |  ||  || ||  |   | ||      ||| |         |||   | ||   |
Db      11633 GTAGAGAATTAGATGTGTCA---AATAAGTTCGTATTAGGAAATATAGCTCGCCTATCAG 11577

Qy        596 AAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACAT 655
              |   |||| || |     | ||||| || |   |||| |  | ||| |    ||| ||
Db      11576 ATCAAAAAAACCAAAGTTATTTATTTAACGTTTTAAAAGAACTCATTTTAATCAAACCA- 11518

Qy        656 CAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTAT 715
                      || ||||||| ||  | | |||  |||| |         | ||||||
Db      11517 -----------AATGTTATTTTACTCCTAGTTGGTAATGGTGAGGATGAGCAGAAATTAA 11469

Qy        716 TGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTG 775
                |    | ||   || | |||||       | ||  | || || |||  ||| | ||
Db      11468 AACAGAAAGCTTTAGAACTAAATCTGACCCCATATGTGCTATTTTTAGGGAGAAGGACTG 11409

Qy        776 ATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTT 835
              |||||   || ||| | | |||    ||| ||||| |  |    ||| | |  || || |
Db      11408 ATATTTCTGATTTATTATCTGCGATGGATGTTTTTTTGCTTCCGTCTAAATATGAGGGGT 11349

Qy        836 TTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTG 895
              |   | |   | | | ||| |||  ||| |  | |    |  || |  |  | ||| | |
Db      11348 TGCCTGTTTCTCTAGTAGAGGCTCAGGCATCGGGATTACAAATTTTATCGTCAGATACAG 11289

Qy        896 GTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTC 955
                  ||| ||  ||        | | |  |    | ||| |||   |  |   |
Db      11288 TGACGCAAGAAGTAGATGTGACCAAAAACATTAGTTACTTACCTATCAACGAAGAGTCTG 11229

Qy        956 TGTTGGCAGAGAAA 969
              |||||  |  ||||
Db      11228 TGTTGCTATGGAAA 11215


RESULT 11
US-10-472-928-4979
; Sequence 4979, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
;  APPLICANT: CHIRON SpA
;  APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
;  TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
;  FILE REFERENCE: P026926WO
;  CURRENT APPLICATION NUMBER: US/10/472,928
;  CURRENT FILING DATE:  2003-09-26
;  PRIOR APPLICATION NUMBER: GB-0107658.7
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 4979
;  SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4979
;   LENGTH: 2162598
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-10-472-928-4979

  Query Match             5.8%;  Score 65.6;  DB 8;  Length 2162598;
  Best Local Similarity   47.8%;  Pred. No. 0.084;
  Matches  265;  Conservative    0;  Mismatches  274;  Indels   15;  Gaps    2;

Qy        416 AAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGA 475
              |||| | | | |   |||    | |   || ||         |    ||||| | |   |
Db     221067 AAGATTCGGCTAAGTGGTTATATGGAGAGAAAGCGTATAAAGACGGAAAAATTGAAATTA 221126

Qy        476 TTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGC 535
              ||   |||   ||| ||   || || | | | |||       |||| ||| | |||   |
Db     221127 TTTTTAATGCAATTGATTTGAAAAAGTATCAATTTTTGTCAGATGTTAGAGAAAAATGTC 221186

Qy        536 GAGATGCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTG 595
              |    |  ||  || ||  |   | ||      ||| |         |||   | ||   |
Db     221187 GTAGAGAATTAGATGTGTCA---AATAAGTTCGTATTAGGAAATATAGCTCGCCTATCAG 221243

Qy        596 AAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACAT 655
              |   |||| || |     | ||||| || |   |||| |  | ||| |    ||| ||
Db     221244 ATCAAAAAAACCAAAGTTATTTATTTAACGTTTTAAAAGAACTCATTTTAATCAAACCA- 221302

Qy        656 CAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTAT 715
                      || ||||||| ||  | | |||  |||| |         | ||||||
Db     221303 -----------AATGTTATTTTACTCCTAGTTGGTAATGGTGAGGATGAGCAGAAATTAA 221351

Qy        716 TGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTG 775
                |    | ||   || | |||||       | ||  | || || |||  ||| | ||
Db     221352 AACAGAAAGCTTTAGAACTAAATCTGACCCCATATGTGCTATTTTTAGGGAGAAGGACTG 221411

Qy        776 ATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTT 835
              |||||   || ||| | | |||    ||| ||||| |  |    ||| | |  || || |
Db     221412 ATATTTCTGATTTATTATCTGCGATGGATGTTTTTTTGCTTCCGTCTAAATATGAGGGGT 221471

Qy        836 TTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTG 895
              |   | |   | | | ||| |||  ||| |  | |    |  || |  |  | ||| | |
Db     221472 TGCCTGTTTCTCTAGTAGAGGCTCAGGCATCGGGATTACAAATTTTATCGTCAGATACAG 221531

Qy        896 GTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTC 955
                  ||| ||  ||        | | |  |    | ||| |||   |  |   |
Db     221532 TGACGCAAGAAGTAGATGTGACCAAAAACATTAGTTACTTACCTATCAACGAAGAGTCTG 221591

Qy        956 TGTTGGCAGAGAAA 969
              |||||  |  ||||
Db     221592 TGTTGCTATGGAAA 221605
```

RESULT 12
US-10-795-159-409/c
; Sequence 409, Application US/10795159
; Publication No. US20050221439A1
; GENERAL INFORMATION:
; APPLICANT: BAKALETZ et al.
; TITLE OF INVENTION: GENES OF AN OTITIS MEDIA ISOLATE OF NONTYPEABLE H. INFLUENZAE
; FILE REFERENCE: 28335/38815A
; CURRENT APPLICATION NUMBER: US/10/795,159
; CURRENT FILING DATE: 2004-03-05
; PRIOR APPLICATION NUMBER: US 60/453,134
; PRIOR FILING DATE: 2003-03-06
; NUMBER OF SEQ ID NOS: 771
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 409
; LENGTH: 3262
; TYPE: DNA
; ORGANISM: H. influenzae
US-10-795-159-409

Query Match 5.7%; Score 63.6; DB 9; Length 3262;
Best Local Similarity 48.7%; Pred. No. 0.014;
Matches 231; Conservative 0; Mismatches 239; Indels 4; Gaps 2;

```
Qy       653 CATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAAT 712
Db       991 CAAAAAAATATCCAGATTGGAATTTGAAAATAGCGGGTTCAGGCGAGGAAGAAGAAAATT 932

Qy       713 TATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAA 772
Db       931 TAAAAAATCTTGCTAAAGCATTGAATATTGAAGATTCAGTTAATTTTATTCCTCGTACTA 872

Qy       773 GTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAG 832
Db       871 ATGATGTTGCTTTTTATTATGAAAGCAGTTCAATTTATTGTTTACCTTCACAAACAGAAG 812

Qy       833 GTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGC---TACCG 889
Db       811 GGTTGCCTTTAGTCGTTATTGAAGCGATGGCATTTGGCTTACCTATTGTTGCCTTTAATT 752

Qy       890 ATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGAT-GTTATCCCTGTCAGTAAT 948
Db       751 GTTCTCCTGGCGTAAAGCAATTAGTCGAACATAAAGAAAATGGTTTTCTTTGTGAACAAA 692

Qy       949 CATATTCTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAA 1008
Db       691 ATAATATTGAAGAAATGGTTAAAGGTTTAGATTTATTAATCAATAATCCTGAACTTTATC 632

Qy      1009 ATAATAGGTATGAAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTG 1068
Db       631 AACAAATGTCAGACAAATCTCGTTTAATGAGTGAAGATTATGGCATTGAGAAAATTATTG 572

Qy      1069 AGTGAGTGGGAGCGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGAT 1122
Db       571 AAGAATGGAAAGGGATTTTATAAGGAATAATTTTATGTTAAAAAAATATTTGAT 518
```

RESULT 13
US-10-795-159-685/c
; Sequence 685, Application US/10795159
; Publication No. US20050221439A1
; GENERAL INFORMATION:
; APPLICANT: BAKALETZ et al.
; TITLE OF INVENTION: GENES OF AN OTITIS MEDIA ISOLATE OF NONTYPEABLE H. INFLUENZAE
; FILE REFERENCE: 28335/38815A
; CURRENT APPLICATION NUMBER: US/10/795,159
; CURRENT FILING DATE: 2004-03-05
; PRIOR APPLICATION NUMBER: US 60/453,134
; PRIOR FILING DATE: 2003-03-06
; NUMBER OF SEQ ID NOS: 771
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 685
; LENGTH: 908766
; TYPE: DNA
; ORGANISM: H. influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9223)..(9223)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (39640)..(39640)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (292404)..(292404)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (292495)..(292495)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (908604)..(908641)
; OTHER INFORMATION: n is a, c, g, or t
US-10-795-159-685

Query Match 5.7%; Score 63.6; DB 9; Length 908766;
Best Local Similarity 48.7%; Pred. No. 0.15;
Matches 231; Conservative 0; Mismatches 239; Indels 4; Gaps 2;

```
Qy       653 CATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAAT 712
Db    727238 CAAAAAAATATCCAGATTGGAATTTGAAAATAGCGGGTTCAGGCGAGGAAGAAGAAAATT 727179

Qy       713 TATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAA 772
Db    727178 TAAAAAATCTTGCTAAAGCATTGAATATTGAAGATTCAGTTAATTTTATTCCTCGTACTA 727119

Qy       773 GTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAG 832
Db    727118 ATGATGTTGCTTTTTATTATGAAAGCAGTTCAATTTATTGTTTACCTTCACAAACAGAAG 727059

Qy       833 GTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGC---TACCG 889
Db    727058 GGTTGCCTTTAGTCGTTATTGAAGCGATGGCATTTGGCTTACCTATTGTTGCCTTTAATT 726999

Qy       890 ATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGAT-GTTATCCCTGTCAGTAAT 948
Db    726998 GTTCTCCTGGCGTAAAGCAATTAGTCGAACATAAAGAAAATGGTTTTCTTTGTGAACAAA 726939

Qy       949 CATATTCTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAA 1008
Db    726938 ATAATATTGAAGAAATGGTTAAAGGTTTAGATTTATTAATCAATAATCCTGAACTTTATC 726879

Qy      1009 ATAATAGGTATGAAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTG 1068
Db    726878 AACAAATGTCAGACAAATCTCGTTTAATGAGTGAAGATTATGGCATTGAGAAAATTATTG 726819

Qy      1069 AGTGAGTGGGAGCGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGAT 1122
Db    726818 AAGAATGGAAAGGGATTTTATAAGGAATAATTTTATGTTAAAAAAATATTTGAT 726765
```

RESULT 14
US-10-282-122A-16045
; Sequence 16045, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith

```
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16045
;   LENGTH: 1092
;   TYPE: DNA
;   ORGANISM: Clostridium botulinum
US-10-282-122A-16045

  Query Match             5.6%;  Score 62.6;  DB 7;  Length 1092;
  Best Local Similarity   46.9%;  Pred. No. 0.014;
  Matches  311;  Conservative    0;  Mismatches  334;  Indels   18;  Gaps    3;

Qy        442 AGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCGAATTTTATTAATACAAATAAA 501
              | || | ||||   ||||   |    || |||||      | | || || || ||||| |
Db        406 ATAATGTCTACCAGTAAAGTAATGGCAGAAGAGACAAATCAATATACTACTAAAAATATA 465

Qy        502 TTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGATGCTTTTAATTTGAAAGACAGT 561
                    | ||   || | |   | ||     | |||        ||| |
Db        466 GAAATAACCCCTTTTGGAGTAAATATAGATACATTTAAGCCTTGCGCCGATAAAATATGAA 525

Qy        562 ACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCAAAAGACTATCCGAACTTATTA 621
              | |  || |    | | | |||| | |   ||| || ||  |||   |   |     ||
Db        526 AAAAAAGAAAATTTAGTAATAGGTACA---GTTAAAACATTAGAGCCTAAATATGGTATA 582

Qy        622 AATGCAATAAATCATTTGATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTT 681
              |     ||  |        ||      ||   |||   |||| || |||  ||
Db        583 GAATACTTAGTTAGAGCTTTTGCAAAGGTAAAGCAAAAACATAATAATATTAAGTTAGAA 642

Qy        682 ATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTT 741
              ||||| || |  || |    ||   |    ||| |  || |    |  | || | ||| |
Db        643 ATTGCAGGAGTAGGGGACCAAAAGGACTTTTTACTTAATCTATGTAATGAACTTAATATA 702

Qy        742 GTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAATTA------------ 789
                 ||  | |||   || || ||    | || | |    || | | ||
Db        703 AAAGACCATGTTAAATTTTTAGGATTTATAAATCAAGAAAAGGTAATAGAGGCTTTTAAT 762

Qy        790 ATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCGTTGTT 849
              |  | || || || | | ||| ||  |  |  ||  | ||| |||| ||  | |  |
Db        763 AGATTTGATGTAGCTGTGTTTCCTTCCACATTAGATAGTGAAAGTTTCGGGGTAGCAGCA 822

Qy        850 GCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAGTTAAAGAA 909
              | ||||||   ||| ||||    ||| ||  ||| | |  ||   || ||| |    |||

Db        823 GTAGAAGCACAGGCATGTGGGACTCCAGTAATTGTTTCTAATGTAGGAGGATTACCTGAA 882

Qy        910 GTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCT---GTTGGCAGAG 966
              |      |  ||| |||||  |  |||     || | |||  | | | |       |  ||
Db        883 GCTACAGCTCCTAATAATAGTAGTATATTAGTAAATAAAAAAAGTGTAGATGAGCTAGCA 942

Qy        967 AAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATGAAAAAT 1026
              |    |   || | | | |   | ||||| ||    ||| | ||  || |||| || |
Db        943 GAGGCCATAGAAAAATTAATTAAGGATGACAATTTAAGGATAAACATGGGTAAGACAGGA 1002

Qy       1027 AGAGAATATATTGTTTCCAATTTTTCAATTAAAACGATAGTGAGTGAGTGGGAGCGCTTA 1086
              ||| ||| | | |     ||||||   ||  ||   |   | | | |    ||     ||
Db       1003 AGAAAATTTGTGGAAGATAATTTTAATATAGAAGATAATTTTAATAATGTAGATATAATA 1062

Qy       1087 TAT 1089
              |||
Db       1063 TAT 1065


RESULT 15
US-10-369-493-34143
; Sequence 34143, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 34143
;   LENGTH: 1101
;   TYPE: DNA
;   ORGANISM: Cytophaga hutchinsonii
US-10-369-493-34143

  Query Match             5.4%;  Score 60.6;  DB 6;  Length 1101;
  Best Local Similarity   53.1%;  Pred. No. 0.036;
  Matches  129;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy        669 TTTTATTTTGCTTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGATTTGGTTTG 728
              || ||  || ||  |||  || |||||| |      |   |     |   | ||    |
Db        684 TTCTAAATTACTGCTTGTAGGTGATGGCCCCGAGCGGGTTCTTATTGAAGGTATGTGCAG 743

Qy        729 TCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTAAAGAATT 788
                |  |    ||    |  | ||   ||| || ||  ||        |   |  |||||
Db        744 AGAGCTTGGCATTTGCAACGATGTGCGCTTTTTAGGAAAACTGGACGCCGTGGAAGAAGT 803

Qy        789 AATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTCTCGTTGT 848
                | | ||  || ||||| ||| || ||  ||||||    ||| |||| ||| | |  |
Db        804 GCTATCTGTAGCCGATCTCTTTATTATGCCTTCTGAAAAAGAAAGTTTCGGTTTAGCCGC 863

Qy        849 TGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAGTTAAAGA 908
                ||||| ||||| ||||||  ||| ||| || |  |  || | || ||| |    ||
Db        864 ACTGGAAGCAATGGCTTGTGAAGTTCCGGTTATTTCATCAAATGCCGGCGGATTGCCGGA 923

Qy        909 AGT 911
              | |
Db        924 ATT 926


Search completed: May 11, 2006, 07:09:40
Job time : 755.533 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:23:58 ; Search time 581.345 Seconds
(without alignments)
7892.621 Million cell updates/sec

Title: US-09-423-093B-1_COPY_11821_12945
Perfect score: 1125
Sequence: 1 atgaagattgtttatataat.....gtaataatataattgattga 1125

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		60.2	5.4	1095	17	US-11-098-686-9848	Sequence 9848, Ap
c 2		60.2	5.4	1457619	17	US-11-098-686-8739	Sequence 8739, Ap
3		51.4	4.6	2020	7	US-09-925-065A-728360	Sequence 728360,
4		50	4.4	1122	17	US-11-098-686-9746	Sequence 9746, Ap
5		50	4.4	1457619	17	US-11-098-686-8739	Sequence 8739, Ap
6		48.6	4.3	173602	17	US-11-121-086-25	Sequence 25, Appl
7		48.4	4.3	1564	11	US-10-932-182A-166165	Sequence 166165,
8		48.4	4.3	1564	11	US-10-932-182A-166165	Sequence 166165,
c 9		48.4	4.3	5026	11	US-10-932-182A-166180	Sequence 166180,
c 10		48.4	4.3	5026	11	US-10-932-182A-166180	Sequence 166180,
11		47.8	4.2	998	12	US-10-301-480-558612	Sequence 558612,
12		47.8	4.2	998	12	US-10-301-480-1172021	Sequence 1172021,
c 13		47.6	4.2	738	12	US-10-301-480-532482	Sequence 532482,
c 14		47.6	4.2	738	12	US-10-301-480-1145891	Sequence 1145891,
c 15		47.2	4.2	819	7	US-09-925-065A-692077	Sequence 692077,
c 16		47.2	4.2	819	7	US-09-925-065A-692078	Sequence 692078,
c 17		45.8	4.1	936	11	US-10-932-182A-166977	Sequence 166977,
c 18		45.8	4.1	936	11	US-10-932-182A-166977	Sequence 166977,
19		45.6	4.1	2526	11	US-10-903-375-34	Sequence 34, Appl
c 20		45.4	4.0	687411	11	US-10-330-773-26	Sequence 26, Appl
c 21		45.2	4.0	613	11	US-10-301-480-33130	Sequence 33130, A
c 22		45.2	4.0	613	12	US-10-301-480-646539	Sequence 646539,
23		45	4.0	244196	10	US-10-995-561-13327	Sequence 13327, A
24		44.6	4.0	621	7	US-09-925-065A-848450	Sequence 848450,
25		44.6	4.0	643	7	US-09-925-065A-526596	Sequence 526596,
26		44.6	4.0	1786	7	US-09-925-065A-694814	Sequence 694814,
27		44.2	3.9	603	12	US-10-301-480-564198	Sequence 564198,
28		44.2	3.9	603	12	US-10-301-480-1177607	Sequence 1177607,
29		44.2	3.9	1269	10	US-10-336-263A-11	Sequence 11, Appl
c 30		44.2	3.9	1368	7	US-09-925-065A-672032	Sequence 672032,
31		44.2	3.9	1392	7	US-09-925-065A-705566	Sequence 705566,
32		44.2	3.9	1392	7	US-09-925-065A-705568	Sequence 705568,
c 33		44.2	3.9	1767	11	US-10-932-182A-6013	Sequence 6013, Ap
c 34		44.2	3.9	1767	11	US-10-932-182A-6013	Sequence 6013, Ap
35		44	3.9	539	7	US-09-925-065A-833868	Sequence 833868,
c 36		44	3.9	1118	7	US-09-925-065A-712432	Sequence 712432,
37		43.8	3.9	1392	7	US-09-925-065A-705567	Sequence 705567,
c 38		43.8	3.9	4116	10	US-10-793-626-4100	Sequence 4100, Ap
39		43.8	3.9	38459	18	US-11-143-564-3	Sequence 3, Appli
40		43.6	3.9	539	7	US-09-925-065A-768139	Sequence 768139,
41		43.6	3.9	1206	7	US-09-925-065A-710881	Sequence 710881,
c 42		43.6	3.9	3660	11	US-10-932-182A-2664	Sequence 2664, Ap
c 43		43.6	3.9	3660	11	US-10-932-182A-2664	Sequence 2664, Ap
44		43.4	3.9	1071	10	US-10-793-626-1753	Sequence 1753, Ap
45		43.4	3.9	3053	10	US-10-793-626-3389	Sequence 3389, Ap

ALIGNMENTS

RESULT 1
US-11-098-686-9848
; Sequence 9848, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9848
; LENGTH: 1095
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-9848

Query Match 5.4%; Score 60.2; DB 17; Length 1095;
Best Local Similarity 45.7%; Pred. No. 0.13;
Matches 209; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

```
Qy        640 ATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCA 699
              ||| |   |||| ||  | ||          |      | | |||||| |||  ||| | |
Db        595 ATTATAAAAAAAGCAAAAGATCTGCCCCCATGGGAAGTACGTATTGCAGGCTCTGGTGAA 654

Qy        700 TTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTC 759
                ||    |   |||| |  ||   |  | | || | ||| | |    | || ||   ||
Db        655 CTATTTCAATCATTACTTAATGAAGCATTTGAACTTAATGTAGCTTCTCAACTTGCTTTA 714

Qy        760 TTGGGGCAAAGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGT 819
              || || |||  |  |  |      |   |||     |||||  | | || ||||
```

```
Db        715 TTAGGTCAACAAGATAGTCGAGTAATTTTACCTCTTGCTGATGCTGTTGTTGTACCCTCA 774

Qy        820 TCTGAGTGGGAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTT 879

Db        775 GTAGAGGGTGAGGGAAGTAATGCAACAATAAAAGAAGGTTGGGCAACAGGAATTCCTGTT 834

Qy        880 GTTGCTACCGATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCT 939

Db        835 ATTGCTTCTAATTTAGAATCAAATGCTGAACTTATTCAAGATGAAATTAATGGACTTCTT 894

Qy        940 GTCAGTAATCATATTCTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAAC 999

Db        895 GTTCCAGTAAATGATCCAGTTGCACTAGCTATGGCTATGATCCGCTGTATGCAAGAACCA 954

Qy       1000 GCAAGAAAAATAATAGGTATGAAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAA 1059

Db        955 CGACTTTGTGAAAAAATTGTATCTGAAGGGACAAAAAGCCTTGAAAATTTTACAACTACA 1014

Qy       1060 ACGATAGTGAGTGAGTGGGAGCGCTTATATTTTAAAT 1096

Db       1015 ACAATGGTAAAGCAATATCTACAACTATATCGTAACT 1051


RESULT 2
US-11-098-686-8739/c
; Sequence 8739, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
;  APPLICANT: Kapur, Vivek and Gebhart, Connie J.
;  TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
;  TITLE OF INVENTION:  FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
;  FILE REFERENCE: 09531-128001
;  CURRENT APPLICATION NUMBER: US/11/098,686
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: PCT/US03/31318
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: US 60/416,395
;  PRIOR FILING DATE: 2002-10-04
;  NUMBER OF SEQ ID NOS: 11433
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8739
;   LENGTH: 1457619
;   TYPE: DNA
;   ORGANISM: Lawsonia intracellularis
US-11-098-686-8739

  Query Match             5.4%;  Score 60.2;  DB 17;  Length 1457619;
  Best Local Similarity   45.7%;  Pred. No. 0.5;
  Matches  209;  Conservative    0;  Mismatches  248;  Indels    0;  Gaps    0;

Qy        640 ATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGGCGCA 699

Db    1185066 ATTATAAAAAAAGCAAAAGATCTGCCCCCATGGGAAGTACGTATTGCAGGCTCTGGTGAA 1185007

Qy        700 TTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTC 759

Db    1185006 CTATTTCAATCATTACTTAATGAAGCATTTGAACTTAATGTAGCTTCTCAACTTGCTTTA 1184947

Qy        760 TTGGGGCAAAGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGT 819

Db    1184946 TTAGGTCAACAAGATAGTCGAGTAATTTTACCTCTTGCTGATGCTGTTGTTGTACCCTCA 1184887

Qy        820 TCTGAGTGGGAAGGTTTTGGTCTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTT 879

Db    1184886 GTAGAGGGTGAGGGAAGTAATGCAACAATAAAAGAAGGTTGGGCAACAGGAATTCCTGTT 1184827

Qy        880 GTTGCTACCGATTCTGGTGGAGTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCT 939

Db    1184826 ATTGCTTCTAATTTAGAATCAAATGCTGAACTTATTCAAGATGAAATTAATGGACTTCTT 1184767

Qy        940 GTCAGTAATCATATTCTGTTGGCAGAGAAAATCGCTGAGACACTTAAAATAGATGATAAC 999

Db    1184766 GTTCCAGTAAATGATCCAGTTGCACTAGCTATGGCTATGATCCGCTGTATGCAAGAACCA 1184707

Qy       1000 GCAAGAAAAATAATAGGTATGAAAAATAGAGAATATATTGTTTCCAATTTTTCAATTAAA 1059

Db    1184706 CGACTTTGTGAAAAAATTGTATCTGAAGGGACAAAAAGCCTTGAAAATTTTACAACTACA 1184647

Qy       1060 ACGATAGTGAGTGAGTGGGAGCGCTTATATTTTAAAT 1096

Db    1184646 ACAATGGTAAAGCAATATCTACAACTATATCGTAACT 1184610


RESULT 3
US-09-925-065A-728360
; Sequence 728360, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 728360
;   LENGTH: 2020
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-728360

  Query Match             4.6%;  Score 51.4;  DB 7;  Length 2020;
  Best Local Similarity   56.1%;  Pred. No. 5.7;
  Matches  97;  Conservative    0;  Mismatches  76;  Indels    0;  Gaps    0;

Qy         95 ATGTTAATATTATTTGTCTAACTGGTATATCTGAGGTAAAGCCAACACAAAATATTAATA 154

Db        878 ATATATATATTAAATATATATATAAAATATATTATATATATTTAAATATATATATAAAAT 937

Qy        155 TTCATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAA 214

Db        938 ATATATATTTAAATATATATATAAAATATATATAATATATAATATAATAATATAATATAA 997

Qy        215 AAAAAATAATTGTCGCCTTAAAGCCAGATATAATACATAGTCATATGTTTCAT 267

Db        998 AAATAATAATTATATATATAATATATCATATATTATATAAATATATATATAAT 1050


RESULT 4
US-11-098-686-9746
; Sequence 9746, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
;  APPLICANT: Kapur, Vivek and Gebhart, Connie J.
;  TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
;  TITLE OF INVENTION:  FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
;  FILE REFERENCE: 09531-128001
;  CURRENT APPLICATION NUMBER: US/11/098,686
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: PCT/US03/31318
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: US 60/416,395
;  PRIOR FILING DATE: 2002-10-04
;  NUMBER OF SEQ ID NOS: 11433
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

; SEQ ID NO 9746
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-9746

Query Match 4.4%; Score 50; DB 17; Length 1122;
Best Local Similarity 50.0%; Pred. No. 9.2;
Matches 125; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```
Qy        722 TGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTA 781
              | |||  |  ||| ||  ||  ||||| |||| | ||  ||||| ||  |   |||||
Db        746 TTGTTCATTTATTAAAAGTTCAGGATAGAGTTGTATTTATGGGGAAAGTAGAAGATATGC 805

Qy        782 AAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTC 841
                |||     | ||   |||| | |||| | |     || | |  || | || |
Db        806 TTTCATTGTATAGAGCAATAGATTTATTTGCTCTTCCCACTTATTATGATGCTTGTTCAA 865

Qy        842 TCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAG 901
              | |||   |||||| | ||| |||| |      ||| ||  ||   | | || |||
Db        866 ATGCTGTATTAGAAGCAAGGGCATGTGGATTAAAAGTTCTTAGTAGTAAATATGATGGTT 925

Qy        902 TTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTGG 961
              ||    | |  |    | |||       |||       ||||||  |       | |
Db        926 CTAGCTCATTTTTACCCCAACATTGGTTAATTAATAACCCTAGTAATTCTCAAGAAATAG 985

Qy        962 CAGAGAAAAT 971
              || | ||| |
Db        986 CAAATAAACT 995
```

RESULT 5
US-11-098-686-8739
; Sequence 8739, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8739
; LENGTH: 1457619
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-8739

Query Match 4.4%; Score 50; DB 17; Length 1457619;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 125; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```
Qy        722 TGGTTTGTCAATTGAATCTTGTGGATAAAGTTTTCTTCTTGGGGCAAAGAAGTGATATTA 781
              | |||  |  ||| ||  ||  ||||| |||| | ||  ||||| ||  |   |||||
Db    1054733 TTGTTCATTTATTAAAAGTTCAGGATAGAGTTGTATTTATGGGGAAAGTAGAAGATATGC 1054792

Qy        782 AAGAATTAATGTGTGCTGCAGATCTTTTTGTTTTGAGTTCTGAGTGGGAAGGTTTTGGTC 841
                |||     | ||   |||| | |||| | |     || | |  || | || |
Db    1054793 TTTCATTGTATAGAGCAATAGATTTATTTGCTCTTCCCACTTATTATGATGCTTGTTCAA 1054852

Qy        842 TCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGGAG 901
              | |||   |||||| | ||| |||| |      ||| ||  ||   | | || |||
Db    1054853 ATGCTGTATTAGAAGCAAGGGCATGTGGATTAAAAGTTCTTAGTAGTAAATATGATGGTT 1054912

Qy        902 TTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTGG 961
              ||    | |  |    | |||       |||       ||||||  |       | |
Db    1054913 CTAGCTCATTTTTACCCCAACATTGGTTAATTAATAACCCTAGTAATTCTCAAGAAATAG 1054972

Qy        962 CAGAGAAAAT 971
              || | ||| |
Db    1054973 CAAATAAACT 1054982
```

RESULT 6
US-11-121-086-25
; Sequence 25, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; LENGTH: 173602
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-25

Query Match 4.3%; Score 48.6; DB 17; Length 173602;
Best Local Similarity 43.6%; Pred. No. 42;
Matches 314; Conservative 0; Mismatches 404; Indels 3; Gaps 2;

```
Qy         99 TAATATTATTTGTCTAACTGGTATATCTGAGGTAAAGCCAACACAAAATAT-TAATATTC 157
              | |||| |||| | ||     ||||  | |  ||       | | || ||||  |||||
Db    136307 TTATATAATTTATATATTATATATAAATTATATATTTTATATATAATATATACAATATAT 136366

Qy        158 ATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAA 217
              || |  |  ||||     | || | || |    || |       | |||   |  || | |
Db    136367 ATAAATTATATATAATATATAAATATATATAAATTATATATAATATATAAAATATATATA 136426

Qy        218 AAATAATTGTCGCCTTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTT 277
              || ||  | |    |  ||   | ||||||   |||   || | | | |    ||   | |
Db    136427 AATTATATATAATATATAAAATATATATAAATTATATAAATCTATATAAATATATACTAT 136486

Qy        278 TTAGTCGTTTTATTAGGATGCTGATTCCAGCGGTGCCCCTGATATGTACCGCACACAACA 337
              |  |   | || ||     |   |  |     |       || | ||  | | | |  |
Db    136487 ATTATAAATATAGTATATATTTATATATATTTATATAATATATTTATATAGTATATATTA 136546

Qy        338 AAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTA 397
              | ||  |  ||   ||| |   ||   |   ||      | | |   ||| ||  || ||
Db    136547 TATATTTATATGTTTATGTATATATAAATATATAATATATAAATATATTTATATATTATA 136606

Qy        398 TTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTA 457
              |  || | |       || | |   |||  | | |||      || | |   || |  |
Db    136607 TATTTATATAATATTTTATATAATATGTTATATATTTTTATATAATATATTATATATATT 136666

Qy        458 AAAATAAAATAGTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTTTGATATTA 517
              | ||   ||| || | | |       | ||| |       || || | || |  |||||
Db    136667 TATATGCTATATTATATAGTATATTATATATAATATATTTTATATATTATATATATATTT 136726

Qy        518 ATGTCAGAAAGAAAACGCGAGATGCTTTTAATTTGAAAGACAGTACAGCAGT--ACTGCT 575
              || | | |   | |         | | | |  |     | | | ||    | |    | |
Db    136727 ATATAATATTTATATATTATATTTATATAATATGTTTATATATTATTATATTTATATAAT 136786

Qy        576 CGCAGTAGGAAGACTTGTTGAAGCAAAAGACTATCCGAACTTATTAAATGCAATAAATCA 635
              |  |    || |     | || |   |    | |||||| ||  |   |||
Db    136787 ATGTTTATATATTATTATATTTATATAATATGTTTATATATTATTATATTTATATAATAT 136846

Qy        636 TTTGATTCTTTCAAAAACATCAAATTGTAATGATTTTATTTTGCTTATTGCTGGCGATGG 695
              || ||  | |  | || ||    | ||||   ||||| |    |||  |
Db    136847 GTTTATAATATATATAATATATTTATATAATATGTTTATATATTATATATATAATATATT 136906

Qy        696 CGCATTAAGAAATAAATTATTGGATTTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTT 755
              || |     | |  ||||  || |    | |   |  |   |  |  ||| | |||
```

```
Db     136907 TATATAATATGTTTATATATTATATATATAATATATTTATATATTATTATATATATTTTA 136966

Qy        756 CTTCTTGGGGCAAAGAAGTGATATTAAAGAATTAATGTGTGCTGCAGATCTTTTTGTTTT 815

Db     136967 TAATATATAATATATATAATATAATATATATTATATATATTATATATATTATATATTA 137026

Qy        816 G 816

Db     137027 G 137027
```

RESULT 7
US-10-932-182A-166165
; Sequence 166165, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166165
; LENGTH: 1564
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166165

Query Match 4.3%; Score 48.4; DB 11; Length 1564;
Best Local Similarity 45.5%; Pred. No. 19;
Matches 212; Conservative 0; Mismatches 251; Indels 3; Gaps 1;

```
Qy         75 AATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGTATATCTGAGGTAAA 134

Db        542 AATAAAAAAACAAGGTATTTATTATATTGGTAATATAATGACAAATAAATTTTATATTGG 601

Qy        135 GCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTT 194

Db        602 TTCAGCAGGATATAATAATATATATGATAGATTTAATAATCATTTATATAGATTACGTGG 661

Qy        195 TAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCAGATATAATACATAG 254

Db        662 TAGTTCTAATATTGCATTAGATGTAAAAAAATATGGTTTAAATACTTTTGTATATGATAT 721

Qy        255 TCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTGATTCCAGCGGTGCC 314

Db        722 AAAAAAAATATATGATAATGAAAGTATTGATAGATATATTTTATATAGTATTGAACAAAT 781

Qy        315 CCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCG 374

Db        782 TTATTTATACCTAAAATTACCTACATATAATAGTGGTAATTTAAAGAATTTTTCTTTATC 841

Qy        375 ACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTT 434

Db        842 TAAATTTGATACTTTAGATCTTATTACTAATAAATTA---AAATATGCTATTACTAATAA 898

Qy        435 TATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCGAATTTTATTAATAC 494

Db        899 TATTTATGGTATTTATTTAGATAAATATATAGCTATAGATAAATCTAATATAATAAAATA 958

Qy        495 AAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540

Db        959 TATAAATAAACATAATAATATTTATTATAATAATTTAAATAATGAT 1004
```

RESULT 8
US-10-932-182A-166165
; Sequence 166165, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166165
; LENGTH: 1564
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166165

Query Match 4.3%; Score 48.4; DB 11; Length 1564;
Best Local Similarity 45.5%; Pred. No. 19;
Matches 212; Conservative 0; Mismatches 251; Indels 3; Gaps 1;

```
Qy         75 AATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGTATATCTGAGGTAAA 134

Db        542 AATAAAAAAACAAGGTATTTATTATATTGGTAATATAATGACAAATAAATTTTATATTGG 601

Qy        135 GCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTT 194

Db        602 TTCAGCAGGATATAATAATATATATGATAGATTTAATAATCATTTATATAGATTACGTGG 661

Qy        195 TAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCAGATATAATACATAG 254

Db        662 TAGTTCTAATATTGCATTAGATGTAAAAAAATATGGTTTAAATACTTTTGTATATGATAT 721

Qy        255 TCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTGATTCCAGCGGTGCC 314

Db        722 AAAAAAAATATATGATAATGAAAGTATTGATAGATATATTTTATATAGTATTGAACAAAT 781

Qy        315 CCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCG 374

Db        782 TTATTTATACCTAAAATTACCTACATATAATAGTGGTAATTTAAAGAATTTTTCTTTATC 841

Qy        375 ACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTT 434

Db        842 TAAATTTGATACTTTAGATCTTATTACTAATAAATTA---AAATATGCTATTACTAATAA 898

Qy        435 TATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCGAATTTTATTAATAC 494

Db        899 TATTTATGGTATTTATTTAGATAAATATATAGCTATAGATAAATCTAATATAATAAAATA 958

Qy        495 AAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540

Db        959 TATAAATAAACATAATAATATTTATTATAATAATTTAAATAATGAT 1004
```

RESULT 9
US-10-932-182A-166180/c
; Sequence 166180, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166180

```
;   LENGTH: 5026
;   TYPE: DNA
;   ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166180

  Query Match             4.3%;  Score 48.4;  DB 11;  Length 5026;
  Best Local Similarity   45.5%;  Pred. No. 24;
  Matches  212;  Conservative    0;  Mismatches  251;  Indels    3;  Gaps    1;

Qy          75 AATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGTATATCTGAGGTAAA 134
               |||    | ||  ||   | ||  || |  || |    | ||   || || | |  |
Db        4998 AATAAAAAAACAAGGTATTTATTATATTGGTAATATAATGACAAATAAATTTTATATTGG 4939

Qy         135 GCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTT 194
                 || ||  | ||| ||||||  || ||     ||   || |    | ||||     |
Db        4938 TTCAGCAGGATATAATAATATATATGATAGATTTAATAATCATTTATATAGATTACGTGG 4879

Qy         195 TAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCAGATATAATACATAG 254
               |||  ||     ||  | || |   || ||      |  |||||  |   | ||    |||
Db        4878 TAGTTCTAATATTGCATTAGATGTAAAAAAATATGGTTTAAATACTTTTGTATATGATAT 4819

Qy         255 TCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTGATTCCAGCGGTGCC 314
                 | |   |  ||| ||||     || |  |     ||   | |   |  |
Db        4818 AAAAAAAATATATGATAATGAAAGTATTGATAGATATATTTTATATAGTATTGAACAAAT 4759

Qy         315 CCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCG 374
               |||       |      | | || |  |||| |||  || ||  |||| ||
Db        4758 TTATTTATACCTAAAATTACCTACATATAATAGTGGTAATTTAAAGAATTTTTCTTTATC 4699

Qy         375 ACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTT 434
                |||| |||||  |  |||||||| ||  ||   ||| | ||| ||    |
Db        4698 TAAATTTGATACTTTAGATCTTATTACTAATAAATTA---AAATATGCTATTACTAATAA 4642

Qy         435 TATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCGAATTTTATTAATAC 494
               |||    | |    |  |  |||| ||| |    |||| |   | ||| | || ||
Db        4641 TATTTATGGTATTTATTTAGATAAATATATAGCTATAGATAAATCTAATATAATAAAATA 4582

Qy         495 AAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540
                |  ||     ||  | ||||| ||   |  ||   ||     |||
Db        4581 TATAAATAAACATAATAATATTTATTATAATAATTTAAATAATGAT 4536


RESULT 10
US-10-932-182A-166180/c
; Sequence 166180, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAO, YOSHIHIRO
;  APPLICANT:  NAKAMURA, NORIHISA
;  APPLICANT:  KODAMA, YUKIKO
;  APPLICANT:  FUJIMURA, TOMOKO
;  APPLICANT:  ASHIKARI, TOSHIHIKO
;  TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
;  FILE REFERENCE: 030685-043
;  CURRENT APPLICATION NUMBER: US/10/932,182A
;  CURRENT FILING DATE:  2004-09-02
;  NUMBER OF SEQ ID NOS: 197023
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166180
;   LENGTH: 5026
;   TYPE: DNA
;   ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166180

  Query Match             4.3%;  Score 48.4;  DB 11;  Length 5026;
  Best Local Similarity   45.5%;  Pred. No. 24;
  Matches  212;  Conservative    0;  Mismatches  251;  Indels    3;  Gaps    1;

Qy          75 AATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGTATATCTGAGGTAAA 134
               |||    | ||  ||   | ||  || |  || |    | ||   || || | |  |
Db        4998 AATAAAAAAACAAGGTATTTATTATATTGGTAATATAATGACAAATAAATTTTATATTGG 4939

Qy         135 GCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTT 194
                 || ||  | ||| ||||||  || ||     ||   || |    | ||||     |
Db        4938 TTCAGCAGGATATAATAATATATATGATAGATTTAATAATCATTTATATAGATTACGTGG 4879

Qy         195 TAGAGCTTTATTTCAAGTAAAAAAAATAATTGTCGCCTTAAAGCCAGATATAATACATAG 254
               |||  ||     ||  | || |   || ||      |  |||||  |   | ||    |||
Db        4878 TAGTTCTAATATTGCATTAGATGTAAAAAAATATGGTTTAAATACTTTTGTATATGATAT 4819

Qy         255 TCATATGTTTCATGCTAATATTTTTAGTCGTTTTATTAGGATGCTGATTCCAGCGGTGCC 314
                 | |   |  ||| ||||     || |  |     ||   | |   |  |
Db        4818 AAAAAAAATATATGATAATGAAAGTATTGATAGATATATTTTATATAGTATTGAACAAAT 4759

Qy         315 CCTGATATGTACCGCACACAACAAAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCG 374
               |||       |      | | || |  |||| |||  || ||  |||| ||
Db        4758 TTATTTATACCTAAAATTACCTACATATAATAGTGGTAATTTAAAGAATTTTTCTTTATC 4699

Qy         375 ACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTT 434
                |||| |||||  |  |||||||| ||  ||   ||| | ||| ||    |
Db        4698 TAAATTTGATACTTTAGATCTTATTACTAATAAATTA---AAATATGCTATTACTAATAA 4642

Qy         435 TATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCGAATTTTATTAATAC 494
               |||    | |    |  |  |||| ||| |    |||| |   | ||| | || ||
Db        4641 TATTTATGGTATTTATTTAGATAAATATATAGCTATAGATAAATCTAATATAATAAAATA 4582

Qy         495 AAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 540
                |  ||     ||  | ||||| ||   |  ||   ||     |||
Db        4581 TATAAATAAACATAATAATATTTATTATAATAATTTAAATAATGAT 4536


RESULT 11
US-10-301-480-558612
; Sequence 558612, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 558612
;   LENGTH: 998
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-558612

  Query Match             4.2%;  Score 47.8;  DB 12;  Length 998;
  Best Local Similarity   47.4%;  Pred. No. 23;
  Matches  210;  Conservative    0;  Mismatches  227;  Indels    6;  Gaps    2;

Qy          90 GCATGATGTTAATATTATTTGTCTAACTGGTATATCTGAGGTAAAGCCAACACAAAATAT 149
               | || ||  |  | ||||   | ||  |    ||| | |  || |      | | |  ||
Db         424 GGATAATTATTCTTTTATAACTTTATATAAATTATTTTATTTATATATTTTATATATGAT 483

Qy         150 TAATATTCATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCA 209
               |  ||||   ||  | ||||| ||  || || |||      | | ||  |    ||| | |
Db         484 TTTTATTTTATAAATAAATATTGAATAATATATTATTTAATATATTTAAATAATATATAA 543

Qy         210 AGTAAAAAAAAATAATTGTCGCCTTAAAGCCAGATATAATACATAGTCATATGTTTCATGC 269
               |    | | | || || |      || |  || ||| ||  ||| | || | ||| |||
Db         544 AAATTATATATTATTTATATTTATATAAACAATTATTATGGATAATTAT-TCTTTTATGA 602

Qy         270 TAATATTTTTAGTCGTTTTATTAGGATGCTGATTCCAGCGGTGCCCCTGATATGTACCGC 329
               |||  |     || ||  ||   |  ||   | ||   | ||     |||||||
Db         603 CTTTATATACCAAGGTCTTTATACCTTTATGTACCAAGGTCTACC-----TATGTACTTT 657

Qy         330 ACACAACAAAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTT 389
               | | |   || |      |||  |   |    | | |  || | |   | |  ||| ||
```

```
Db        658 AGAGAGTCAATAGTTTCTTGGATAATTATTCTTTTATACTTTTACATAAATTATTTTATT 717

Qy        390 AGCTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGC 449

Db        718 TATATATTTTATATATTATTTATATTTTATGAATAATAAATATTCAATAATATACTATTT 777

Qy        450 TACACCTAAAAATAAAATAGTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTT 509

Db        778 AATATATTTATATAATATATATTATGTATTACATTTAAATAATGTATATATTATTTATAT 837

Qy        510 TGATATTAATGTCAGAAAGAAAA 532

Db        838 TTATATAAATATAATTATGGATA 860
```

RESULT 12
US-10-301-480-1172021
; Sequence 1172021, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1172021
; LENGTH: 998
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-1172021

Query Match 4.2%; Score 47.8; DB 12; Length 998;
Best Local Similarity 47.4%; Pred. No. 23;
Matches 210; Conservative 0; Mismatches 227; Indels 6; Gaps 2;

```
Qy         90 GCATGATGTTAATATTATTTGTCTAACTGGTATATCTGAGGTAAAGCCAACACAAAATAT 149

Db        424 GGATAATTATTCTTTTATAACTTTATATAAATTATTTTATTTATATATTTTATATATGAT 483

Qy        150 TAATATTCATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCA 209

Db        484 TTTTATTTTATAAATAAATATTGAATAATATATTATTTAATATATTTAAATAATATATAA 543

Qy        210 AGTAAAAAAAAATAATTGTCGCCTTAAAGCCAGATATAATACATAGTCATATGTTTCATGC 269

Db        544 AAATTATATATTATTTATTATATTTATATAAACAATTATTATGGATAATTAT-TCTTTTATGA 602

Qy        270 TAATATTTTTAGTCGTTTTATTAGGATGCTGATTCCAGCGGTGCCCCTGATATGTACCGC 329

Db        603 CTTTATATACCAAGGTCTTTATACCTTTATGTACCAAGGTCTACC-----TATGTACTTT 657

Qy        330 ACACAACAAAAATGAAGGTGGCAATGCAAGGATGTTTTGTTATCGACTGAGTGATTTTTT 389

Db        658 AGAGAGTCAATAGTTTCTTGGATAATTATTCTTTTATACTTTTACATAAATTATTTTATT 717

Qy        390 AGCTTCTATTACTACAAATGTAAGTAAAGAGGCTGTTCAAGAGTTTATAGCAAGAAAGGC 449

Db        718 TATATATTTTATATATTATTTATATTTTATGAATAATAAATATTCAATAATATACTATTT 777

Qy        450 TACACCTAAAAATAAAATAGTAGAGATTCCGAATTTTATTAATACAAATAAATTTGATTT 509

Db        778 AATATATTTATATAATATATATTATGTATTACATTTAAATAATGTATATATTATTTATAT 837

Qy        510 TGATATTAATGTCAGAAAGAAAA 532

Db        838 TTATATAAATATAATTATGGATA 860
```

RESULT 13
US-10-301-480-532482/c
; Sequence 532482, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 532482
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-532482

Query Match 4.2%; Score 47.6; DB 12; Length 738;
Best Local Similarity 51.9%; Pred. No. 23;
Matches 107; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

```
Qy         98 TTAATATTATTTGTCTAACTGGTATATCTGAGGTAAAGCCAACACAAAATATTAATATTC 157

Db        604 TTAAATATATATTTATATATAATATATTATAAATATAATATATATTTATTATGTTATATA 545

Qy        158 ATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAA 217

Db        544 AATATATTTATATTTATAATATATTACATAGATTTATATATTATATATTTATTATATAAT 485

Qy        218 AAATAATTGTCGCCTTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTT 277

Db        484 ATATAGATTTATACATAAATATATATAACATATGTATTTATATATATATATATTTTTT 425

Qy        278 TTAGTCGTTTTATTAGGATGCTGATT 303

Db        424 TTTTTTTTTTTTTGGATACGGAGTTT 399
```

RESULT 14
US-10-301-480-1145891/c
; Sequence 1145891, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1145891
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-1145891

Query Match 4.2%; Score 47.6; DB 12; Length 738;
Best Local Similarity 51.9%; Pred. No. 23;
Matches 107; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

```
Qy          98 TTAATATTATTTGTCTAACTGGTATATCTGAGGTAAAGCCAACACAAAATATTAATATTC 157
               ||||   ||| | | ||  |  |||||  |  || |    | |   | |||       |
Db         604 TTAAATATATATTTATATATAATATATTATAAATATAATATATATTTATTATGTTATATA 545

Qy         158 ATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAA 217
               | ||| || ||||  |||| |  ||    || ||| |      | |||||    || ||
Db         544 AATATATTTATATTTATAATATATTACATAGATTTATATATTATATATTTATTATATAAT 485

Qy         218 AAATAATTGTCGCCTTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTT 277
               | |||  | |   | ||||   | |||  |||  || | |||| | |     || | |||
Db         484 ATATAGATTTATACATAAATATATATAACATATGTATTTATATATATATATATATTTTTT 425

Qy         278 TTAGTCGTTTTATTAGGATGCTGATT 303
               ||  |  |||| |    | |  | ||
Db         424 TTTTTTTTTTTTTGGATACGGAGTTT 399


RESULT 15
US-09-925-065A-692077/c
; Sequence 692077, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 692077
;   LENGTH: 819
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-692077

  Query Match             4.2%;  Score 47.2;  DB 7;  Length 819;
  Best Local Similarity   51.5%;  Pred. No. 28;
  Matches  106;  Conservative    1;  Mismatches   99;  Indels    0;  Gaps    0;

Qy          98 TTAATATTATTTGTCTAACTGGTATATCTGAGGTAAAGCCAACACAAAATATTAATATTC 157
               ||||   ||| | | ||  |  |||||   |  || |    | |   | |||       |
Db         634 TTAAATATATATTTATATATAATATATTATAAATATAATATATATTTATTATGTTATATA 575

Qy         158 ATTATGTTAATATGGATAAAAATTTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAA 217
               | ||| || ||||  |||| |  ||    || ||| |      | |||||    || ||
Db         574 AATATATTTATATTTATAATATATTACATAGATTTATATATTATATATTTATTATATAAT 515

Qy         218 AAATAATTGTCGCCTTAAAGCCAGATATAATACATAGTCATATGTTTCATGCTAATATTT 277
               | |||  | |   | ||||   | |||  |||  || | |||| | |     || | :||
Db         514 ATATAGATTTATACATAAATATATATAACATATGTATTTATATATATATATATTTWTT 455

Qy         278 TTAGTCGTTTTATTAGGATGCTGATT 303
               ||  |  |||| |    | |  | ||
Db         454 TTTTTTTTTTTTTGGATACGGAGTTT 429


Search completed: May 11, 2006, 08:30:45
Job time : 586.345 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:45:22 ; Search time 2842.13 Seconds
(without alignments)
15660.244 Million cell updates/sec

Title: US-09-423-093B-2_COPY_79_861
Perfect score: 783
Sequence: 1 atgaacaaagaaaccgtttc.....gaaaggagactaaaaagtga 783

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	783	100.0	14002	1	AF061251	AF061251 Escherici
2	783	100.0	14024	6	BD085555	Nucleic a
3	783	100.0	14029	6	AX029262	AX029262 Sequence
c 4	781.4	99.8	14181	6	BD184789	BD184789 Nucleic a
c 5	781.4	99.8	14187	6	AR204225	AR204225 Sequence
c 6	781.4	99.8	14187	6	AR637574	AR637574 Sequence
7	781.4	99.8	31482	1	AB008676	AB008676 Escherich
c 8	781.4	99.8	110000	1	AE005174_28	Continuation (29 o
c 9	781.4	99.8	110000	1	BA000007_27	Continuation (28 o
10	360.6	46.1	14323	1	AY730593	AY730593 Salmonell
11	321.6	41.1	14224	1	AY730592	AY730592 Citrobact
c 12	140.4	17.9	4915	6	BD193531	BD193531 Enterococ
c 13	140.4	17.9	302241	1	AE016953	AE016953 Enterococ
14	133.4	17.0	110000	1	AE017355_15	Continuation (16 o
15	129.8	16.6	110000	1	AE017225_15	Continuation (16 o
16	129.8	16.6	110000	1	AE017334_15	Continuation (16 o
17	129.8	16.6	290525	1	AE017029	AE017029 Bacillus
18	124.8	15.9	110000	1	AP008934_00	AP008934 Staphyloc
19	123.8	15.8	347456	1	AE017003	AE017003 Bacillus
c 20	122.6	15.7	110000	1	CR626927_43	Continuation (44 o
c 21	121	15.5	321052	1	AE017281	AE017281 Bacillus
22	116.8	14.9	9901	1	U32768	U32768 Haemophilus
23	116.8	14.9	110000	6	BD426631_09	Continuation (10 o
24	116.8	14.9	110000	6	AR274513_09	Continuation (10 o
25	116.8	14.9	110000	6	AR632719_09	Continuation (10 o
26	113	14.4	10911	1	AF402315	AF402315 Shigella
c 27	112.4	14.4	110000	1	AP006627_32	Continuation (33 o
28	110	14.0	110000	1	AE017355_51	Continuation (52 o
29	107	13.7	110000	1	CP000083_21	Continuation (22 o
30	100	12.8	891	6	AR481796	AR481796 Sequence
31	100	12.8	891	6	AR527395	AR527395 Sequence
32	100	12.8	891	6	AR609674	AR609674 Sequence
33	100	12.8	891	6	AX934453	AX934453 Sequence
c 34	100	12.8	11442	1	AF400048	AF400048 Campyloba
c 35	99.2	12.7	110000	1	BA000004_37	Continuation (38 o
c 36	96.4	12.3	24425	1	AY422197	AY422197 Campyloba
c 37	95.8	12.2	110000	1	BA000028_30	Continuation (31 o
38	95.2	12.2	3738	1	AF164515	AF164515 Streptoco
39	95	12.1	276034	1	AE017015	AE017015 Bacillus
c 40	94.8	12.1	11474	1	AF215659	AF215659 Campyloba
c 41	94.8	12.1	12390	1	AF401528	AF401528 Campyloba
c 42	94.8	12.1	24437	1	AF167344	AF167344 Campyloba
c 43	94.6	12.1	11455	1	AY044868	AY044868 Campyloba
c 44	93.6	12.0	12370	1	AY422196	AY422196 Campyloba
45	93.2	11.9	906	6	BD249797	BD249797 Campyloba

ALIGNMENTS

RESULT 1
AF061251
LOCUS AF061251 14002 bp DNA linear BCT 20-AUG-1998
DEFINITION Escherichia coli serotype O157:H7 O antigen gene cluster.
ACCESSION AF061251
VERSION AF061251.1 GI:3435170
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 14002)
AUTHORS Wang,L. and Reeves,P.R.
TITLE Organization of Escherichia coli O157 O antigen gene cluster and identification of its specific genes
JOURNAL Infect. Immun. 66 (8), 3545-3551 (1998)
PUBMED 9673232
REFERENCE 2 (bases 1 to 14002)
AUTHORS Wang,L.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1998) Microbiology, The University of Sydney, NSW 2006, Australia
FEATURES Location/Qualifiers
source 1. .14002
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="C664-1992"
/serotype="O157:H7"
/db_xref="taxon:562"
gene 60. .842
/gene="wbdN"
CDS 60. .842
/gene="wbdN"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdN"
/protein_id="AAC32339.1"
/db_xref="GI:3435171"
/translation="MNKETVSIIMPVYNGAKTIISSVESIIHQSYQDFVLYIIDDCST
DDTFSLINSRYKNNQKIRILRNKTNLGVAESRNYGIEMATGKYISFCDADDLWHEKKL
ERQIEVLNNECVDVVCSNYYVIDNNRNIVGEVNAPHVINYRKMLMKNYIGNLTGIYNA

```
                NKLGKFYQKKIGHEDYLMWLEIINKTNGAICIQDNLAYYMRSNNSLSGNKIKAAKWTW
                SIYREHLHLSFPKTLYYFLLYASNGVMKKITHSLLRRKETKK"
gene            839. .2023
                /gene="wzy"
CDS             839. .2023
                /gene="wzy"
                /codon_start=1
                /transl_table=11
                /product="O antigen polymerase Wzy"
                /protein_id="AAC32340.1"
                /db_xref="GI:3435172"
                /translation="MKSAAKLIFLFLFTLYSLQLYGVIIDDRITNFDTKVLTSIIIIF
                QIFFVLLFYLTIINERKQQKKFIVNWELKLILVFLFVTIEIAAVVLFLKEGIPIFDDD
                PGGAKLRIAEGNGLYIRYIKYFGNIVVFALIILYDEHKFKQRTIIFVYFTTIALFGYR
                SELVLLILQYILITNILSKDNRNPKIKRIIGYFLLVGVVCSLFYLSLGQDGEQNDSYN
                NMLRIINRLTIEQVEGVPYVVSESIKNDFFPTPELEKELKAIINRIQGIKHQDLFYGE
                RLHKQVFGDMGANFLSVTTYGAELLVFFGFLCVFIIPLGIYIPFYLLKRMKKTHSSIN
                CAFYSYIIMILLQYLVAGNASAFFFGPFLSVLIMCTPLILLHDTLKRLSRNENISYNC
                DL"
gene            1992. .2738
                /gene="wbdO"
CDS             1992. .2738
                /gene="wbdO"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase WbdO"
                /protein_id="AAC32341.1"
                /db_xref="GI:3435173"
                /translation="MKISVITVTYNNAEGLEKTLSSLSILKIKPFEIIIVDGGSTDGT
                NRVISRFTSMNITHVYEKDEGIYDAMNKGRMLAKGDLIHYLNAGDSVIGDIYKNIKEP
                CLIKVGLFENDKLLGFSSITHSNTGYCHQGVIFPKNHSEYDLRYKICADYKLIQEVFP
                EGLRSLSLITSGYVKYDMGGVSSKKRILRDKELAKIMFEKNKKNLIKFIPISIIKILF
                PERLRRVLRKMQYICLTLFFMKNSSPYDNE"
gene            2725. .4116
                /gene="wzx"
CDS             2725. .4116
                /gene="wzx"
                /codon_start=1
                /transl_table=11
                /product="O antigen flippase Wzx"
                /protein_id="AAC32342.1"
                /db_xref="GI:3435174"
                /translation="MIMNKIKKILKFCTLKKYDTSSALGREQERYRIISLSVISSLIS
                KILSLLSLILTVSLTLPYLGQERFGVWMTITSLGAALTFLDLGIGNALTNRIAHSFAC
                GKNLKMSRQISGGLTLLAGLSFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKV
                FVIIFGLGIYSNGVQKVYMGIQKAYISNIVNAIFILLSIITLVISSKLHAGLPVLIVS
                TLGIQYISGIYLTINLIIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSG
                DNFIISITLGVTYVAVFSITQRLFQISTVPLTIYNIPLWAAYADAHARNDTQFIKKTL
                RTSLKIVGISSFLLAFILVVFGSEVVNIWTEGKIQVPRTFIIAYALWSVIDAFSNTFA
                SFLNGLNIVKQQMLAVVTLILIAIPAKYIIVSHFGLTVMLYCFIFIYIVNYFIWYKCS
                FKKHIDRQLNIRG"
gene            4113. .5213
                /gene="per"
CDS             4113. .5213
                /gene="per"
                /codon_start=1
                /transl_table=11
                /product="perosamine synthetase Per"
                /protein_id="AAC32343.1"
                /db_xref="GI:3435175"
                /translation="MKMKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKF
                AEQNHVQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFV
                DSDNETWQMSVSDIEQKITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEA
                FGSKYKGKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHR
                QYWHDVIGYNYRMTNICAAIGLAQLEQADDFISRKREIADIYKKNINSLVQVHKESKD
                VFHTYWMVSILTRTAEEREELRNHLADKLIETRPVFYPVHTMPMYSEKYQKHPIAEDL
                GWRGINLPSFPSLSNEQVIYICESINEFYSDK"
gene            5238. .6452
                /gene="wbdP"
CDS             5238. .6452
                /gene="wbdP"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase WbdP"
                /protein_id="AAC32344.1"
                /db_xref="GI:3435176"
                /translation="MKIALNSDGFYEWGGGIDFIKYILSILETKPEICIDILLPRNDI
                HSLIREKAFPFKSILKAILKRERPRWISLNRFNEQYYRDAFTQNNIETNLTFIKSKSS
                AFYSYFDSSDCDVILPCMRVPSGNLNKKAWIGYIYDFQHCYYPSFFSKREIDQRNVFF
                KLMLNCANNIIVNAHSVITDANKYVGNYSAKLHSLPFSPCPQLKWFADYSGNIAKYNI
                DKDYFIICNQFWKHKDHATAFRAFKIYTEYNPDVYLVCTGATQDYRFPGYFNELMVLA
                KKLGIESKIKILGHIPKLEQIELIKNCIAVIQPTLFEGGPGGGVTFDAIALGKKVILS
                DIDVNKEVNCGDVYFFQAKNHYSLNDAMVKADESKIFYEPTTLIELGLKRRNACADFL
                LDVVKQEIESRS"
gene            6472. .7590
                /gene="gmd"
CDS             6472. .7590
                /gene="gmd"
                /codon_start=1
                /transl_table=11
                /product="GDP-D-mannose dehydratase Gmd"
                /protein_id="AAC32345.1"
                /db_xref="GI:3435177"
                /translation="MTKVALITGVTGQDGSYLAEFLLDKGYEVHGIKRRASSFNTERI
                DHIYQDPHGSNPNFHLHYGDLTDSSNLTRILKEVQPDEVYNLAAMSHVAVSFESPEYT
                ADVDAIGTLRLLEAIRFLGLENKTRFYQASTSELYGLVQEIPQKESTPFYPRSPYAVA
                KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRGLANIAQGLESCLYL
                GNMDSLRDWGHAKDYVRMQWLMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLGIKMSF
                VGKGIEEKGIVDSVEGQDAPGVKPGDVIVAVDPRYFRPAEVDTLLGDPSKANLKLGWR
                PEITLAEMISEMVAKDLEAAKKHSLLKSHGFSVSLALE"
gene            7587. .8558
                /gene="fcl"
CDS             7587. .8558
                /gene="fcl"
                /codon_start=1
                /transl_table=11
                /product="fucose synthetase Fcl"
                /protein_id="AAC32346.1"
                /db_xref="GI:3435178"
                /translation="MMMNKQRIFIAGHQGMVGSAITRRLKQRDDVELVLRTRDELNLL
                DSSAVLDFFSSQKIDQVYLAAAKVGGILANSSYPADFIYENIMIEANVIHAAHKNNVN
                KLLFLGSSCIYPKLAHQPIMEDELLQGKLEPTNEPYAIAKIAGIKLCESYNRQFGRDY
                RSVMPTNLYGPNDNFHPSNSHVIPALLRRFHDAVENNSPNVVVWGSGTPKREFLHVDD
                MASASIYVMEMPYDIWQKNTKVMLSHINIGTGIDCTICELAETIAKVVGYKGHITFDT
                TKPDGAPRKLLDVTLLHQLGWNHKITLHKGLENTYNWFLENQLQYRG"
gene            8561. .9070
                /gene="wbdQ"
CDS             8561. .9070
                /gene="wbdQ"
                /codon_start=1
                /transl_table=11
                /product="WbdQ"
                /protein_id="AAC32347.1"
                /db_xref="GI:3435179"
                /translation="MFLHSQDFATIVRSTPLISIDLIVENEFGEILLGKRINRPAQGY
                WFVPGGRVLKDEKLQTAFERLTEIELGIRLPLSVGKFYGIWQHFYEDNSMGGDFSTHY
                IVIAFLLKLQPNILKLPKSQHNAYCWLSRAKLINDDDVHYNCRAYFNNKTNDAIGLDN
                KDIICLMRQ"
gene            9052. .10500
                /gene="manC"
CDS             9052. .10500
                /gene="manC"
                /codon_start=1
                /transl_table=11
                /product="mannose-1-P guanosyltransferase ManC"
                /protein_id="AAC32348.1"
                /db_xref="GI:3435180"
                /translation="MSDAPIIAVVMAGGTGSRLWPLSRELYPKQFLQLSGDNTLLQTT
                LLRLSGLSCQKPLVITNEQHRFVVAEQLREINKLNGNIILEPCGRNTAPAIAISAFHA
                LKRNPQEDPLLLVLAADHVIAKESVFCDAIKNATPIANQGKIVTFGIIPEYAETGYGY
                IERGELSVPLQGHENTGFYYVNKFVEKPNRETAELYMTSGNHYWNSGIFMFKASVYLE
                ELRKFRPDIYNVCEQVASSSYIDLDFIRLSKEQFQDCPAESIDFAVMEKTEKCVVCPV
                DIGWSDVGSWQSLWDISLKSKTGDVCKGDILTYDTKNNYIYSESALVAAIGIEDMVIV
                QTKDAVLVSKKSDVQHVKKIVEMLKLQQRTEYISHREVFRPWGKFDSIDQGERYKVKK
                IIVKPGEGLSLRMHHHRSEHWIVLSGTAKVTLGDKTKLVTANESIYIPLGAAYSLENP
                GIIPLNLIEVSSGDYLGEDDIIRQKERYKHED"
```

```
     gene            10504. .11874
                     /gene="manB"
     CDS             10504. .11874
                     /gene="manB"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphomannomutase ManB"
                     /protein_id="AAC32349.1"
                     /db_xref="GI:3435181"
                     /translation="MKSLTCFKAYDIRGKLGEELNEDIAWRIGRAYGEFLKPKTIVLG
                     GDVRLTSEALKLALAKGLQDAGVDVLDIGMSGTEEIYFATFHLGVDGGIEVTASHNPM
                     DYNGMKLVREGARPISGDTGLRDVQRLAEANDFPPVDETKRGRYQQINLRDAYVDHLF
                     GYINVKNLTPLKLVINSGNGAAGPVVDAIEARFKALGAPVELIKVHNTPDGNFPNGIP
                     NPLLPECRDDTRNAVIKHGADMGIAFDGDFDRCFLFDEKGQFIEGYYIVGLLAEAFLE
                     KNPGAKIIHDPRLSWNTVDVVTAAGGTPVMSKTGHAFIKERMRKEDAIYGGEMSAHHY

  Query Match             100.0%;  Score 783;  DB 1;  Length 14002;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-111;
  Matches  783;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 119

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 179

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 239

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 299

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 359

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 419

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 479

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 539

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 599

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 659

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 719

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 779

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 839

Qy        781 TGA 783
              |||
Db        840 TGA 842


RESULT 2
BD085555
LOCUS       BD085555               14024 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085555
VERSION     BD085555.1  GI:22631165
KEYWORDS    JP 2001522248-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 14024)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 2 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Escherichia coli
            PN   JP 2001522248-A/2
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14024
            FT                   /organism='Escherichia coli'.
FEATURES             Location/Qualifiers
     source          1. .14024
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 783;  DB 6;  Length 14024;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-111;
  Matches  783;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 138

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 198

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 258

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 318

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 378

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 438

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 498

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        499 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 558

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 618

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 678

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 738

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 798

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 858

Qy        781 TGA 783
              |||
Db        859 TGA 861
```

RESULT 3
AX029262

```
LOCUS       AX029262               14029 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 2 from Patent WO9850531.
ACCESSION   AX029262
VERSION     AX029262.1  GI:10190113
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 2 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .14029
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
ORIGIN
```

```
  Query Match           100.0%;  Score 783;  DB 6;  Length 14029;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-111;
  Matches  783;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 138

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 198

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 258

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 318

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 378

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 438

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 498

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 558

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 618

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 678

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 738

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 798

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 858

Qy        781 TGA 783
              |||
Db        859 TGA 861
```

RESULT 4
BD184789/c

```
LOCUS       BD184789               14181 bp    DNA     linear   PAT 17-JUN-2003
DEFINITION  Nucleic acid molecule and polypeptide specific to intestinal
            hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
            thereof.
ACCESSION   BD184789
VERSION     BD184789.1  GI:31876989
KEYWORDS    JP 2002355074-A/25.
SOURCE      Escherichia coli O157:H7
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 14181)
  AUTHORS   Hayashi,H., Shinagawa,H., Makino,K., Hayashi,T., Onishi,S.,
            Hattori,M. and Kurokawa,K.
  TITLE     Nucleic acid molecule and polypeptide specific to intestinal
            hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
  JOURNAL   Patent: JP 2002355074-A 25 10-DEC-2002;
            PRESIDENT OF UNIVERSITY OF TSUKUBA
COMMENT     OS   Escherichia coli O157:H7
            PN   JP 2002355074-A/25
            PD   10-DEC-2002
            PF   24-JAN-2002 JP 2002015959
            PI   HIDEO HAYASHI,HIDEO SHINAGAWA,KOZO MAKINO,TETSUYA HAYASHI,SHIN

            PI   ONISHI,
            PI   MASAHIRA HATTORI,KEN KUROKAWA
            PC   C12N15/09,C12N15/09,A61K31/7088,A61K39/00,A61K48/00,A61P31/04,
            PC   C07K14/245,
            PC   C07K16/12,C12M1/00,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/
            PC   02,C12Q1/68,
            PC   G01N33/15,G01N33/50,G01N33/53,G01N33/53,G01N33/566,G01N37/00,
            PC   C12N15/00,
```

```
            PC   C12N15/00,C12N5/00
            CC   Nucleic acid molecule and polypeptide specific to intestinal
            CC        hemorrhagic
            CC   pathogenic Escherichia coli O157:H7, and method of use thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14181
            FT                   /organism='Escherichia coli O157:H7'.
FEATURES             Location/Qualifiers
     source          1. .14181
                     /organism="Escherichia coli O157:H7"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:83334"
ORIGIN

  Query Match             99.8%;  Score 781.4;  DB 6;  Length 14181;
  Best Local Similarity   99.9%;  Pred. No. 2.6e-111;
  Matches  782;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13917 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 13858

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13857 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 13798

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13797 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 13738

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13737 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 13678

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13677 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 13618

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13617 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 13558

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      13557 TATTATGTTATAGATAACAAAAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 13498

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13497 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 13438

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13437 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 13378

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13377 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 13318

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13317 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 13258

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13257 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 13198

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13197 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 13138

Qy        781 TGA 783
              |||
Db      13137 TGA 13135
```

```
RESULT 5
AR204225/c
LOCUS       AR204225              14187 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 121 from patent US 6365723.
ACCESSION   AR204225
VERSION     AR204225.1  GI:21500820
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14187)
  AUTHORS   Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
  TITLE     Sequences of E. coli O157
  JOURNAL   Patent: US 6365723-A 121 02-APR-2002;
FEATURES             Location/Qualifiers
     source          1. .14187
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             99.8%;  Score 781.4;  DB 6;  Length 14187;
  Best Local Similarity   99.9%;  Pred. No. 2.6e-111;
  Matches  782;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13922 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 13863

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13862 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 13803

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13802 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 13743

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13742 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 13683

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13682 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 13623

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13622 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 13563

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      13562 TATTATGTTATAGATAACAAAAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 13503

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13502 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 13443

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13442 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 13383

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13382 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 13323

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13322 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 13263

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13262 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 13203
```

```
Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13202 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 13143

Qy        781 TGA 783
              |||
Db      13142 TGA 13140


RESULT 6
AR637574/c
LOCUS       AR637574               14187 bp    DNA     linear   PAT 20-APR-2005
DEFINITION  Sequence 121 from patent US 6855814.
ACCESSION   AR637574
VERSION     AR637574.1  GI:62771316
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14187)
  AUTHORS   Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
  TITLE     Sequences of E. coli O157
  JOURNAL   Patent: US 6855814-A 121 15-FEB-2005;
            Wisconsin Alumni Research Foundation; Madison, WI
FEATURES             Location/Qualifiers
     source          1. .14187
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match           99.8%;  Score 781.4;  DB 6;  Length 14187;
  Best Local Similarity   99.9%;  Pred. No. 2.6e-111;
  Matches  782;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13922 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 13863

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13862 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 13803

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13802 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 13743

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13742 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 13683

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13682 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 13623

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13622 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 13563

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      13562 TATTATGTTATAGATAACAAAAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 13503

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13502 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 13443

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13442 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 13383

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13382 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 13323

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13322 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 13263

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13262 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 13203

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13202 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 13143

Qy        781 TGA 783
              |||
Db      13142 TGA 13140


RESULT 7
AB008676
LOCUS       AB008676               31482 bp    DNA     linear   BCT 19-MAY-1999
DEFINITION  Escherichia coli 0157 DNA, map position at 46 min., complete cds.
ACCESSION   AB008676
VERSION     AB008676.1  GI:4867913
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Shimizu,T., Yamasaki,S., Tsukamoto,T. and Takeda,Y.
  TITLE     Analysis of the genes responsible for the O-antigen synthesis in
            enterohaemorrhagic Escherichia coli O157
  JOURNAL   Microb. Pathog. 26 (5), 235-247 (1999)
   PUBMED   10222209
REFERENCE   2  (bases 1 to 31482)
  AUTHORS   Shimizu,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1997) Takeshi Shimizu, International Medical
            Center of Japan, Department of Appropriate Technology Development,
            Research Institute; 1-21-1 Toyama, Shinjuku-ku, Tokyo 162, Japan
            (E-mail:kshimizu@imcj.go.jp, Tel:+81-3-3202-7181,
            Fax:+81-3-3202-7364)
FEATURES             Location/Qualifiers
     source          1. .31482
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="184"
                     /serotype="O157"
                     /db_xref="taxon:562"
                     /map="46 min."
                     /clone="1-4-1"
     gene            505. .1398
                     /gene="galF"
     CDS             505. .1398
                     /gene="galF"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="UTP-glucose-1-phosphate uridylyltransferase"
                     /protein_id="BAA77723.1"
                     /db_xref="GI:4867914"
                     /translation="MTNLKAVIPVAGLGMHMLPATKAIPKEMLPIVDKPMIQYIVDEI
                     VAAGIKEILLVTHASKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNV
                     RQGEPLGLGHSILCARPAIGDNPFVVVLPDVVIDDASADPLRYNLAAMIARFNETGRS
                     QVLAKRMPGDLSEYSVIQTKEPLDREGKVSRIVEFIEKPDQPQTLDSDIMAVGRYVLS
                     ADIWPELERTQPGAWGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKY
                     GLRNLKEGAKFRKGIEKLLSE"
     gene            1717. .2499
                     /gene="wbhA"
     CDS             1717. .2499
                     /gene="wbhA"
```

```
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="glycosyl transferase"
                /protein_id="BAA77724.1"
                /db_xref="GI:4867915"
                /translation="MNKETVSIIMPVYNGAKTIISSVESIIHQSYQDFVLYIIDDCST
                DDTFSLINSRYKNNQKIRILRNKTNLGVAESRNYGIEMATGKYISFCDADDLWHEKKL
                ERQIEVLNNECVDVVCSNYYVIDNKRNIVGEVNAPHVINYRKMLMKNYIGNLTGIYNA
                NKLGKFYQKKIGHEDYLMWLEIINKTNGAICIQDNLAYYMRSNNSLSGNKIKAAKWTW
                SIYREHLHLSFPKTLYYFLLYASNGVMKKITHSLLRRKETKK"
gene            3153. .3680
                /gene="wbhB"
CDS             3153. .3680
                /gene="wbhB"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="NADH dehydrogenase"
                /protein_id="BAA77725.1"
                /db_xref="GI:4867916"
                /translation="MLRIINRLTIEQVESVPYVVSESIKNDFFPTPELEKELKAIINR
                IQGIKHQDLFYGERLHKQVFGDMGANFLSVTTYGAELLVFFGFLCVFIIPLGIYIPFY
                LLKRMKKTHSSINCAFYSYIIMILLQYLVAGNASAFFFGPFLSVLIMCTPLILLHDTL
                KRLSRNENISYNCDL"
gene            3649. .4395
                /gene="wbhC"
CDS             3649. .4395
                /gene="wbhC"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="glycosyl transferase"
                /protein_id="BAA77726.1"
                /db_xref="GI:4867917"
                /translation="MKISVITVTYNNAEGLEKTLSSLSILKIKPFEIIIVDGGSTDGT
                NRVISRFTSMNITHVYEKDEGIYDAMNKGRMLAKGDLIHYLNAGDSVIGDIYKNIKEP
                CLIKVGLFENDKLLGFSSITHSNTGYCHQGVIFPKNHSEYDLRYKICADYKLIQEVFP
                EGLRSLSLITSGYVKYDMGGVSSKKRILRDKELAKIMFEKNKKNLIKFIPISIIKILF
                PERLRRVLRKMQYICLTLFFMKNSSPYDNE"
gene            4388. .5773
                /gene="wzx"
CDS             4388. .5773
                /gene="wzx"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="flippase"
                /protein_id="BAA77727.1"
                /db_xref="GI:4867918"
                /translation="MNKIKKILKFCTLKKYDTSSALGREQERYRIISLSVISSLISKI
                LSLLSLILTVSLTLPYLGQERFGVWMTITSLGAALTFLDLGIGNALTNRIAHSFACGK
                NLKMSRQISGGLTLLAGLSFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKVFV
                IIFGLGIYSNGVQKVYMGIQKAYISNIVNAIFILLSIITLVISSKLHAGLPVLIVSTL
                GIQYISGIYLTINLIIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSGDN
                FIISITLGVTYVAVFSITQRLFQISTVPLTIYNIPLWAAYADAHARNDTQFIKKTLRT
                SLKIVGISSFLLAFILVVFGSEVVNIWTEGKIQVPRTFIIAYALWSVIDAFSNTFASF
                LNGLNIVKQQMLAVVTLILIAIPAKYIIVSHFGLTVMLYCFIFIYIVNYFIWYKCSFK
                KHIDRQLNIRG"
gene            5776. .6870
                /gene="wbhD"
CDS             5776. .6870
                /gene="wbhD"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="perosamine synthetase"
                /protein_id="BAA77728.1"
                /db_xref="GI:4867919"
                /translation="MKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKFAE
                QNHVQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFVDS
                DNETWQMSVSDIEQKITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEAFG
                SKYKGKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHRQY
                WHDVIGYNYRMTNICAAIGLAQLEQADDFISRKREIADIYKKNINSLVQVHKESKDVF
                HTYWMVSILTRTAEERELRNHLADKLIETRPVFYPVHTMPMYSEKYQKHPIAEDLGW
                RGINLPSFPSLSNEQVIYICESINEFYSDK"
gene            6895. .8109
                /gene="wbhE"
CDS             6895. .8109
                /gene="wbhE"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="glycosyl transferase"
                /protein_id="BAA77729.1"
                /db_xref="GI:4867920"
                /translation="MKIALNSDGFYEWGGGIDFIKYILSILETKPEICIDILLPRNDI
                HSLIREKAFPFKSILKAILKRERPRWISLNRFNEQYYRDAFTQNNIETNLTFIKSKSS
                AFYSYFDSSDCDVILPCMRVPSGNLNKKAWIGYIYDFQHCYYPSFFSKREIDQRNVFF
                KLMLNCANNIIVNAHSVITDANKYVGNYSAKLHSLPFSPCPQLKWFADYSGNIAKYNI
                DKDYFIICNQFWKHKDHATAFRAFKIYTEYNPDVYLVCTGATQDYRFPGYFNELMVLA
                KKLGIESKIKILGHIPKLEQIELIKNCIAVIQPTLFEGGPGGGVTFDAIALGKKVILS
                DIDVNKEVNCGDVYFFQAKNHYSLNDAMVKADESKIFYEPTTLIELGLKRRNACADFL
                LDVVKQEIESRS"
gene            8129. .9247
                /gene="gmd"
CDS             8129. .9247
                /gene="gmd"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose 4,6-dehydratase"
                /protein_id="BAA77730.1"
                /db_xref="GI:4867921"
                /translation="MTKVALITGVTGQDGSYLAEFLLDKGYEVHGIKRRASSFNTERI
                DHIYQDPHGSNPNFHLHYGDLTDSSNLTRILKEVQPDEVYNLAAMSHVAVSFESPEYT
                ADVDAIGTLRLLEAIRFLGLENKTRFYQASTSELYGLVQEIPQKESTPFYPRSPYAVA
                KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRGLANIAQGLESCLYL
                GNMDSLRDWGHAKDYVRMQWLMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLGIKMSF
                VGKGIEEKGIVDSVEGQDAPGVKPGDVIVAVDPRYFRPAEVDTLLGDPSKANLKLGWR
                PEITLAEMISEMVAKDLEAAKKHSLLKSHGFSVSLALE"
gene            9250. .10215
                /gene="wbhF"
CDS             9250. .10215
                /gene="wbhF"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="GDP-L-fucose pathway enzyme"
                /protein_id="BAA77731.1"
                /db_xref="GI:4867922"
                /translation="MNKQRIFIAGHQGMVGSAITRRLKQRDDVELVLRTRDELNLLDS
                SAVLDFFSSQKIDQVYLAAAKVGGILANSSYPADFIYENIMIEANVIHAAHKNNVNKL
                LFLGSSCIYPKLAHQPIMEDELLQGKLEPTNEPYAIAKIAGIKLCESYNRQFGRDYRS
                VMPTNLYGPNDNFHPSNSHVIPALLRRFHDAVENNSPNVVVWGSGTPKREFLHVDDMA
                SASIYVMEMPYDIWQKNTKVMLSHINIGTGIDCTICELAETIAKVVGYKGHITFDTTK
                PDGAPRKLLDVTLLHQLGWNHKITLHKGLENTYNWFLENQLQYRG"
gene            10218. .10727
                /gene="wbhG"
CDS             10218. .10727
                /gene="wbhG"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="GDP-L-fucose pathway enzyme"
                /protein_id="BAA77732.1"
                /db_xref="GI:4867923"
                /translation="MFLHSQDFATIVRSTPLISIDLIVENEFGEILLGKRINRPAQGY
                WFVPGGRVLKDEKLQTAFERLTEIELGIRLPLSVGKFYGIWQHFYEDNSMGGDFSTHY
                IVIAFLLKLQPNILKLPKSQHNAYCWLSRAKLINDDDVHYNCRAYFNNKTNDAIGLDN
                KDIICLMRQ"
gene            10739. .12157
                /gene="manC"
CDS             10739. .12157
                /gene="manC"
                /note="putative"
```

```
                /codon_start=1
                /transl_table=11

  Query Match             99.8%;  Score 781.4;  DB 1;  Length 31482;
  Best Local Similarity   99.9%;  Pred. No. 1.9e-111;
  Matches  782;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1717 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 1776

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1777 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 1836

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1837 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 1896

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1897 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 1956

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1957 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 2016

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2017 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 2076

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       2077 TATTATGTTATAGATAACAAAAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 2136

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2137 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 2196

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2197 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 2256

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2257 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 2316

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2317 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 2376

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2377 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 2436

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2437 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 2496

Qy        781 TGA 783
              |||
Db       2497 TGA 2499


RESULT 8
AE005174_28/c
WPCOMMENT
  Sequence split into 56 fragments  LOCUS AE005174 Accession AE005174
     Fragment Name      Begin     End
     AE005174_00            1    110000
     AE005174_01       100001    210000
     AE005174_02       200001    310000
     AE005174_03       300001    410000
     AE005174_04       400001    510000
     AE005174_05       500001    610000
     AE005174_06       600001    710000
     AE005174_07       700001    810000
     AE005174_08       800001    910000
     AE005174_09       900001   1010000
     AE005174_10      1000001   1110000
     AE005174_11      1100001   1210000
     AE005174_12      1200001   1310000
     AE005174_13      1300001   1410000
     AE005174_14      1400001   1510000
     AE005174_15      1500001   1610000
     AE005174_16      1600001   1710000
     AE005174_17      1700001   1810000
     AE005174_18      1800001   1910000
     AE005174_19      1900001   2010000
     AE005174_20      2000001   2110000
     AE005174_21      2100001   2210000
     AE005174_22      2200001   2310000
     AE005174_23      2300001   2410000
     AE005174_24      2400001   2510000
     AE005174_25      2500001   2610000
     AE005174_26      2600001   2710000
     AE005174_27      2700001   2810000
     AE005174_28      2800001   2910000
     AE005174_29      2900001   3010000
     AE005174_30      3000001   3110000
     AE005174_31      3100001   3210000
     AE005174_32      3200001   3310000
     AE005174_33      3300001   3410000
     AE005174_34      3400001   3510000
     AE005174_35      3500001   3610000
     AE005174_36      3600001   3710000
     AE005174_37      3700001   3810000
     AE005174_38      3800001   3910000
     AE005174_39      3900001   4010000
     AE005174_40      4000001   4110000
     AE005174_41      4100001   4210000
     AE005174_42      4200001   4310000
     AE005174_43      4300001   4410000
     AE005174_44      4400001   4510000
     AE005174_45      4500001   4610000
     AE005174_46      4600001   4710000
     AE005174_47      4700001   4810000
     AE005174_48      4800001   4910000
     AE005174_49      4900001   5010000
     AE005174_50      5000001   5110000
     AE005174_51      5100001   5210000
     AE005174_52      5200001   5310000
     AE005174_53      5300001   5410000
     AE005174_54      5400001   5510000
     AE005174_55      5500001   5528445
Continuation (29 of 56) of AE005174 from base 2800001 (AE005174 Escherichia coli O157:H7

  Query Match             99.8%;  Score 781.4;  DB 1;  Length 110000;
  Best Local Similarity   99.9%;  Pred. No. 1.2e-111;
  Matches  782;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57488 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 57429

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57428 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 57369

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57368 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 57309

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57308 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 57249
```

```
Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57248 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 57189

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57188 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 57129

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      57128 TATTATGTTATAGATAACAAAAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 57069

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57068 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 57009

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57008 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 56949

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56948 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 56889

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56888 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 56829

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56828 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 56769

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56768 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 56709

Qy        781 TGA 783
              |||
Db      56708 TGA 56706
```

RESULT 9
BA000007_27/c
WPCOMMENT
Sequence split into 55 fragments LOCUS BA000007 Accession BA000007

Fragment Name	Begin	End
BA000007_00	1	110000
BA000007_01	100001	210000
BA000007_02	200001	310000
BA000007_03	300001	410000
BA000007_04	400001	510000
BA000007_05	500001	610000
BA000007_06	600001	710000
BA000007_07	700001	810000
BA000007_08	800001	910000
BA000007_09	900001	1010000
BA000007_10	1000001	1110000
BA000007_11	1100001	1210000
BA000007_12	1200001	1310000
BA000007_13	1300001	1410000
BA000007_14	1400001	1510000
BA000007_15	1500001	1610000
BA000007_16	1600001	1710000
BA000007_17	1700001	1810000
BA000007_18	1800001	1910000
BA000007_19	1900001	2010000
BA000007_20	2000001	2110000
BA000007_21	2100001	2210000
BA000007_22	2200001	2310000
BA000007_23	2300001	2410000
BA000007_24	2400001	2510000
BA000007_25	2500001	2610000
BA000007_26	2600001	2710000
BA000007_27	2700001	2810000
BA000007_28	2800001	2910000
BA000007_29	2900001	3010000
BA000007_30	3000001	3110000
BA000007_31	3100001	3210000
BA000007_32	3200001	3310000
BA000007_33	3300001	3410000
BA000007_34	3400001	3510000
BA000007_35	3500001	3610000
BA000007_36	3600001	3710000
BA000007_37	3700001	3810000
BA000007_38	3800001	3910000
BA000007_39	3900001	4010000
BA000007_40	4000001	4110000
BA000007_41	4100001	4210000
BA000007_42	4200001	4310000
BA000007_43	4300001	4410000
BA000007_44	4400001	4510000
BA000007_45	4500001	4610000
BA000007_46	4600001	4710000
BA000007_47	4700001	4810000
BA000007_48	4800001	4910000
BA000007_49	4900001	5010000
BA000007_50	5000001	5110000
BA000007_51	5100001	5210000
BA000007_52	5200001	5310000
BA000007_53	5300001	5410000
BA000007_54	5400001	5498450

Continuation (28 of 55) of BA000007 from base 2700001 (BA000007 Escherichia coli O157:H7

Query Match 99.8%; Score 781.4; DB 1; Length 110000;
Best Local Similarity 99.9%; Pred. No. 1.2e-111;
Matches 782; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87274 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 87215

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87214 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 87155

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87154 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 87095

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87094 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 87035

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87034 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 86975

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86974 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 86915

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      86914 TATTATGTTATAGATAACAAAAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 86855

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86854 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 86795

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86794 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 86735

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      86734 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 86675

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86674 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 86615

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86614 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 86555

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86554 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 86495

Qy        781 TGA 783
              |||
Db      86494 TGA 86492
```

RESULT 10
AY730593

```
LOCUS       AY730593               14323 bp    DNA     linear   BCT 25-OCT-2004
DEFINITION  Salmonella enterica subsp. enterica serovar Urbana UDP-glucose
            pyrophosphorylase (galF) gene, partial cds; O-antigen gene cluster,
            complete sequence; and 6-phosphogluconate dehydrogenase (gnd) gene,
            partial cds.
ACCESSION   AY730593 AY012194
VERSION     AY730593.1  GI:54401420
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Urbana
  ORGANISM  Salmonella enterica subsp. enterica serovar Urbana
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 14323)
  AUTHORS   Jensen,S.O. and Reeves,P.R.
  TITLE     Molecular evolution of the GDP-mannose pathway genes (manB and
            manC) in Salmonella enterica
  JOURNAL   Microbiology (Reading, Engl.) 147 (Pt 3), 599-610 (2001)
  PUBMED    11238967
REFERENCE   2  (bases 1 to 14323)
  AUTHORS   Samuel,G., Hogbin,J.P., Wang,L. and Reeves,P.R.
  TITLE     Relationships of the Escherichia coli O157, O111, and O55 O-Antigen
            Gene Clusters with Those of Salmonella enterica and Citrobacter
            freundii, Which Express Identical O Antigens
  JOURNAL   J. Bacteriol. 186 (19), 6536-6543 (2004)
  PUBMED    15375135
REFERENCE   3  (bases 1 to 14323)
  AUTHORS   Jenkins,S.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-2000) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
REFERENCE   4  (bases 1 to 14323)
  AUTHORS   Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2004) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
  REMARK    Sequence update by submitter
COMMENT     On Oct 25, 2004 this sequence version replaced gi:11992707.
FEATURES             Location/Qualifiers
     source          1. .14323
                     /organism="Salmonella enterica subsp. enterica serovar
                     Urbana"
                     /mol_type="genomic DNA"
                     /strain="M284"
                     /serotype="30"
                     /serovar="Urbana"
                     /sub_species="enterica"
                     /db_xref="taxon:260677"
     gene            <1. .852
                     /gene="galF"
     CDS             <1. .852
                     /gene="galF"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-glucose pyrophosphorylase"
                     /protein_id="AAV34505.1"
                     /db_xref="GI:54401421"
                     /translation="MHMLPATKAIPKEMLPIVDKPMIQYIVDEIVAAGIKEIVLVTHA
                     SKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNVRQAQPLGLGHSILC
                     ARPVVGDNPFIVVLPDIIIDDATADPLRYNLAAMVARFNETGRSQVLAKRMKGDLSEY
                     SVIQTKEPLDNEGKVSRIVEFIEKPDQPQTLDSDLMAVGRYVLSADIWAELERTEPGA
                     WGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKYGLRNLKEGAKFRKS
                     IEQLLHE"
     gene            1220. .2002
                     /gene="wbdN"
     CDS             1220. .2002
                     /gene="wbdN"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAV34506.1"
                     /db_xref="GI:54401422"
                     /translation="MNREIVSVIMPVYNGSHTIVDSIESVLNQTYKDIKLYVIDDCST
                     DDTAKLLTVNYANNKNVKIIISSKNKGVAASRNLGIKESKGKYIAFCDADDLWENNKI
                     EKQINLLEHNNINVVCSNYYIMDFNKKIIGETHSPVTISYKKMLMKNYVGNLTGVYNA
                     EKLGKFYQKKVGHEDYLMWLEIIKKSNFAICLQENLAFYRVSNNSLSASKLVAAKWTW
                     NIYRECLGLSLPKSLYYFSLYAANGVIKKIGLINSDNSKAKE"
     gene            2255. .3157
                     /gene="wzy"
     CDS             2255. .3157
                     /gene="wzy"
                     /codon_start=1
                     /transl_table=11
                     /product="O-antigen polymerase"
                     /protein_id="AAV34507.1"
                     /db_xref="GI:54401423"
                     /translation="MEIMAIILFAKEGIPIFDADPGGAKLKIADGNGIYIRYIKYFGN
                     IVTFTLLLLLNKQKVKQKIIAIIYFLTIALFGYRSELVLLFLQYYIIMSIVNEGGNKY
                     KFLNKIFIFIFMGGVIGSLFYFSLGQDGAVNNSTENIDRIINRLTIEQVESVPYIISE
                     SLEYDLFPTPELGKEINAIFNRILGEKDQNLFYGERLHKKVFGDMGANFLSVTTYGAE
                     LLAFFGPLCIFLIPIGLCIPFAMLKKISNTHSTVNGAIYSFLIMILLQYLIAGNASAF
                     FFGPFITAIVMSIPLMILHDILKR"
     gene            3163. .3894
                     /gene="wbdO"
     CDS             3163. .3894
                     /gene="wbdO"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAV34508.1"
                     /db_xref="GI:54401424"
                     /translation="MNTSVITVTYNNAEGLNKTLESLCLLKVKPYEIIIIDANSIDNT
                     KAIVKKYSDMNIIYICEKDNGIYDAMNKGLRLARGKLLHYLNAGDCVSGEIYADLNGP
                     CLLKVGLFDKDNLIGFSSLTHSNTGYCHQGIIFPSKHPIYDLKYKLCADYKLIQMVFP
                     MGLKQLPLISNGFIKYDMGGVSANKRFLRDKELLQIMLETNKKTLIKFIPLMVIKMVF
                     PKKIRRILRRVNYYFSLILTNKKNV"
     gene            3896. .5287
                     /gene="wzx"
     CDS             3896. .5287
                     /gene="wzx"
                     /codon_start=1
                     /transl_table=11
                     /product="O-antigen flippase"
                     /protein_id="AAV34509.1"
                     /db_xref="GI:54401425"
                     /translation="MIKEKIKILFKYCTLRKYNTTTDYGRQQERYRIISLSIISNFIS
                     KLLALLSLILTVSLTLPYLGQDRFGIWMTIASFSAALTFLDLGIGNALTNRIAHSFAR
                     GKKIRMIRQISGGLTLLAILAFIIIAIFLTIANFVNWGSLLKGINDNTHDEVQFAIKL
                     FAIIFGIGIFSNGVQRIYMGMQKAYISYISNAFFILISIISLFISSRLHAGIPMLILS
                     TIGVQYLSGLSLTILLIYKKLITLNKIRINVKKETPYLISNGFFFFILQMGTLATWSG
                     DNFIISTTLGVTYVSILSITQRLFQISTVPLTIYNAPLWAAYADAHARHDVDFIRKTL
                     LGSLKIVGISSTTLALILVIFGQSIIAIWTGEELRVPQSLIVAYAVWSIIDAYSNTLS
                     SFLNGLNIIKQQVFAVVTLILISIPSKYFIVSHFGLTAMLSCFIILYVTNYLIWYKCS
                     FKKSIHLQLNIRG"
     gene            5290. .6384
                     /gene="per"
```

```
CDS             5290. .6384
                /gene="per"
                /codon_start=1
                /transl_table=11
                /product="perosamine synthetase"
                /protein_id="AAV34510.1"
                /db_xref="GI:54401426"
                /translation="MEYIPVYQPALSGKEKEYVNDCIDSTWISSKGSYIQRFEKKFAE
                QNNIQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPVFVDS
                DLQTWQISVDDIKNKITSRTKAIMCVHLYGHPCDMSPILEIAKVNNLFIIEDCAEAFG
                SKYKDKFVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAAHRQY
                WHDVIGYNYRMTNICAAIGLAQLEQAQLFIARKREIAEIYKNNLTPLIEVHQEGNDIF
                HTYWMVSILARNAEEREVLRSFLSEKLTETRPLFYPVHTMPMYSDKYQKHPVAENLGW
                RGINLPSFPTLSNEQVIYICNAINEFYKDK"
gene            6409. .7623
                /gene="wbdP"
CDS             6409. .7623
                /gene="wbdP"
                /codon_start=1
                /transl_table=11
                /product="glycosyltransferase"
                /protein_id="AAV34511.1"
                /db_xref="GI:54401427"
                /translation="MKITLNTDGFNEWGGGVDFIKYLLRTLEIEQDIFLDVLLPREDI
                SSLLKRNIFPVKNMIKAILSGQTPRWKKLEGFNEQYYKHAFAEITEQSKLSFITGKSS
                AYFSHYIKSDYDIILPCMRTPAVNMKNRAWLGYIYDFQHCYYPSFFSEREIAKRNLFF
                KRMLDNANNVIVNSHAVISDAKKFVGDFSANLHALPFSPCPELNWLNDNSDLIDKYAI
                DKDYFIICNQFWKHKDHSTAFKAFKKYIEINPDVYLVCTGATQDYRFPGYFDELMCLL
                KKLGISEKVKILGHIPKLEQIELIKNSIAVIQPTLFEGGPGGGITFDAVALGKKIVLS
                DIEVNKDVDCGHVYFFHSQNEYSLTEALVKATAVDHYYDPHVLLESGIKRRKKCADFI
                LNVISKEIESRI"
gene            7647. .8765
                /gene="gmd"
CDS             7647. .8765
                /gene="gmd"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose dehydratase"
                /protein_id="AAV34512.1"
                /db_xref="GI:54401428"
                /translation="MSKVALITGVTGQDGSYLAEFLLEKGYEVHGIKRRASSFNTERV
                DHIYQDPHSCNPKFHLHYGDLTDASNLTRILQEVQPDEVYNLGAMSHVAVSFESPEYT
                ADVDAMGTLRLLEAIRFLGLEKKTRFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
                KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRAIANIAQGLESCLYL
                GNMDSLRDWGHAKDYVRMQWMMLQQEQPEDFVIATGVQYSVRQFVELAAAQLGIKLRF
                EGEGINEKGIVVSVTGHDAPGVKPGDVIVAVDPRYFRPAEVETLLGDPSKAHEKLGWK
                PEITLSEMVSEMVANDLEAAKKHSLLKSHGFNVNLSLE"
gene            8768. .9733
                /gene="fcl"
CDS             8768. .9733
                /gene="fcl"
                /codon_start=1
                /transl_table=11
                /product="fucose synthetase"
                /protein_id="AAV34513.1"
                /db_xref="GI:54401429"
                /translation="MNKQRIFVAGHRGMVGSAIVRQLAQRGDVELVLRTRDELDLLDG
                RAVQAFFAGTGIDQVYLAAAKVGGIVANNTYPADFIYENMMIESNIIHAAHLHNVNKL
                LFLGSSCIYPKLARQPMAESELLQGTLEPTNEPYAIAKIAGIKLCESYNRQYGRDYRS
                VMPTNLYGPHDNFHPDNSHVIPALLRRFHEAAQSHAPEVVVWGSGTPMREFLHVDDMA
                AASIHVMELSREVWQENTAPMLSHINVGTGVDCTISELAQTIAKVVGYQGRVVFDAAK
                PDGTPRKLLDVTRLHQLGWYHEISLEAGLAGTYQWFLENQQRFRG"
gene            9736. .10236
                /gene="gmm"
CDS             9736. .10236
                /gene="gmm"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose mannosylhydrolase"
                /protein_id="AAV34514.1"
                /db_xref="GI:54401430"
                /translation="MFLRQEDFAAVVRATPLISLDFIVENGQGEILLGQRLNRPAQGY
                WFVPGGRVRKDETLEAAFARLTQAELGVRLPLAAGTFYGVWQHFYDDNFSGEDFSTHY
                IVLGFRLRVAESDLLLPDAQHGSYRWLTPEQLLASDNVHDNSRAYFQNAPYSVIGLDK
                KDVKYV"
```

```
Query Match             46.1%;  Score 360.6;  DB 1;  Length 14323;
Best Local Similarity   67.8%;  Pred. No. 1.3e-46;
Matches  504;  Conservative    0;  Mismatches  239;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||| | |||||  |||||| |||||||||| || || || ||  || | |||||  |
Db       1220 ATGAATAGAGAAATTGTTTCAGTAATTATGCCTGTATATAACGGTTCCCACACTATTGTT 1279

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
                ||  |||||||| || || ||||| | ||  ||||| |    || ||  | || ||
Db       1280 GACTCTATAGAATCAGTTTTAAATCAAACCTACAAAGATATAAAGTTATACGTTATCGAT 1339

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              |||||||| || ||||| ||    |   ||| | | |       ||    |||||| | ||
Db       1340 GATTGTAGTACAGATGACACTGCTAAGTTATTGACCGTAAATTATGCTAACAATAAAAAT 1399

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              |||  ||  |   |   | ||||  ||| || || | |||| |||| |  |||||
Db       1400 GTAAAGATTATAATAAGTTCAAAAAATAAAGGGGTCGCTGCAAGTAGAAACTTAGGAATT 1459

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              |     |||  || || |||||| | |||||||| || |||||| | ||| |  | ||
Db       1460 AAGGAATCCAAAGGAAAGTATATTGCATTTTGTGACGCTGATGATCTTTGGGAAAATAAC 1519

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||| | ||    |||||  |  | ||| |  || |   | |  |||| || ||||| ||
Db       1520 AAAATTGAAAAGCAAATTAATTTATTAGAGCATAATAATATTAATGTTGTTTGTTCCAAC 1579

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              || ||  |||| |||   |||| ||| ||  |||| |||    ||  |||        |||
Db       1580 TACTACATTATGGATTTTAATAAAAAAATAATTGGTGAAACCCATAGTCCAGTCACTATA 1639

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              | ||||| |||||||||| |||||||||||| | || ||| | |||||  | ||||||||
Db       1640 AGTTATAAAAAAATGCTTATGAAAAACTACGTTGGAAATCTAACAGGTGTGTATAATGCA 1699

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              | |||||||||| |||||||||||||  | ||||| || |||||| | |||||||||
Db       1700 GAAAAATTGGGTAAATTTTATCAAAAAAAAGTAGGTCATGAAGATTATCTTATGTGGCTG 1759

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||| || || || |  |||  |||||||||| | ||||| ||||| || | ||| |
Db       1760 GAAATTATAAAAAAGAGCAATTTTGCTATTTGTCTACAAGAAAATCTTGCCTTTTATAGA 1819

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              |||  || |||||  |||| |  | |||| |    || || ||||||||||||| |||
Db       1820 GTTTCGAACAATTCTTTATCCGCAAGTAAATTAGTTGCAGCGAAATGGACATGGAATATT 1879

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              || ||||||  ||||   || ||| | || ||| | ||||||||||| |  ||||||||
Db       1880 TACAGAGAATGTTTAGGGTTATCCCTCCCGAAATCTTTATATTATTTCTCTTTATATGCA 1939

Qy        721 TCAAATGGAGTCATGAAAAAAAT 743
              | |||||||| || ||||||||
Db       1940 GCTAATGGAGTAATAAAAAAAAT 1962
```

```
RESULT 11
AY730592
LOCUS       AY730592               14224 bp    DNA     linear   BCT 02-NOV-2004
DEFINITION  Citrobacter freundii serotype 90 UDP-glucose pyrophosphorylase
            (galF) gene, partial cds; O-antigen gene cluster, complete
            sequence; and 6-phosphogluconate dehydrogenase (gnd) gene, partial
            cds.
ACCESSION   AY730592
VERSION     AY730592.2  GI:55228555
KEYWORDS    .
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
```

```
            Enterobacteriaceae; Citrobacter.
REFERENCE   1  (bases 1 to 14224)
  AUTHORS   Samuel,G., Hogbin,J.P., Wang,L. and Reeves,P.R.
  TITLE     Relationships of the Escherichia coli O157, O111, and O55 O-Antigen
            Gene Clusters with Those of Salmonella enterica and Citrobacter
            freundii, Which Express Identical O Antigens
  JOURNAL   J. Bacteriol. 186 (19), 6536-6543 (2004)
   PUBMED   15375135
REFERENCE   2  (bases 1 to 14224)
  AUTHORS   Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2004) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
REFERENCE   3  (bases 1 to 14224)
  AUTHORS   Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2004) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
  REMARK    Sequence update by submitter
COMMENT     On Nov 2, 2004 this sequence version replaced gi:54401406.
FEATURES             Location/Qualifiers
     source          1. .14224
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /strain="M1972"
                     /serotype="90"
                     /db_xref="taxon:546"
     gene            <1. .773
                     /gene="galF"
     CDS             <1. .773
                     /gene="galF"
                     /codon_start=3
                     /transl_table=11
                     /product="UDP-glucose pyrophosphorylase"
                     /protein_id="AAV34492.2"
                     /db_xref="GI:55228556"
                     /translation="FVIVAAGIKEIVLVTHASKNAVENHFDTSYELESLLEQRVKRQL
                     LAEVQSICPPGVTIMNVRQAQLLGLGHSILCARPIIGDNPFVVVLPDVVLDTASADPL
                     RYNLAAMVARFNETGRSQVLAKRMEGDLSEYSVIQTKDPLDNEGKVSRIVEFIEKPDQ
                     PQTLDSDLMAVGRYVLSADIWAELEKTAPGAWGRIQLTDAISEPAKKQSVDAILMTGE
                     SYDCGKKMGYMQAFVKYGLRNLKEGPKFRKSIEKLLAM"
     gene            1104. .1853
                     /gene="wbdN"
     CDS             1104. .1853
                     /gene="wbdN"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAV34493.1"
                     /db_xref="GI:54401408"
                     /translation="METESVSIIMPVYNGASTIVQSIESILNQTYKDIRLYIIDDCST
                     DETSILLQNLYISNDRVELIRNIKNKGVAESRNIGIEQSRGKYIAFCDADDLWGETKL
                     EKQLDILESEKVDVVCTNYYLIDLEDNIIGSVCSPGIITYKTMLGKNHIGNLTGIYNV
                     KKIGKVYQKKIGHEDYLMWLEILRKTEQAICIQENLAKYRVSKKSLSANKLIAAQWTW
                     NIYRKQLKIPIIPALYFFSQYAVQGLLKKII"
     gene            1890. .3047
                     /gene="wzy"
     CDS             1890. .3047
                     /gene="wzy"
                     /codon_start=1
                     /transl_table=11
                     /product="O-antigen polymerase"
                     /protein_id="AAV34494.1"
                     /db_xref="GI:54401409"
                     /translation="MKKSVGIISFFLFSFYVLQLYGLLIDVRIKDVNIQIIAGGLVLF
                     QILCVLFYALFIRKNLSYQSGLFFNSKKIIIIFLLTFVTIELAAVILFLKEGIPIFDA
                     DPGGAKLRIADGNGIYIRYIKYFGNIVCFSLCLLLKDRKILHYSVVIIYFSTISLFGY
                     RSELIMLMIQYYILMSIINNNKVKHSMLRKVVILFSISCIVIGLFYFSLGQEGQENNY
                     ASNVNRIFNRLTIEQVESVPYIINESMKYDLFPTSEIGKEINAIANRFQGIKDNSLFY
                     GERLHKDIFGDMGENFLSVTTFGAELLAFFGPFCFFLIPLGIYIPYFLLIKMSQTKNT
                     VLCSAYSYSVVIFLQYLIAGNFSAFFFGPILTIVIMCIPLVLLHDVLKRLN"
     gene            3047. .3781
                     /gene="wbdO"
     CDS             3047. .3781
                     /gene="wbdO"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAV34495.1"
                     /db_xref="GI:54401410"
                     /translation="MKISVVTVTYNNGNGLRKTLQSLSSIKTKPCEIIIIDAESKDNT
                     KSVVDEYFSKMNIIYINEKDNGIYDAMNKGHKIANGDYIHYLNAGDCVSGDVYANVTG
                     TCLIKVGLFNERDSFLGFSNLTHSNTGYCHQGVIFPKEHLDYDLKYKLCADYDLLQKT
                     FPEGLASLPLITDGYIKYDMSGVSANKTLLRDKELLAIMWKNNKRIIPKFMIYVLFKT
                     MIPKGCRRFLRKLKYSISVFSLNGNK"
     gene            3789. .5180
                     /gene="wzx"
     CDS             3789. .5180
                     /gene="wzx"
                     /codon_start=1
                     /transl_table=11
                     /product="O-antigen flippase"
                     /protein_id="AAV34496.1"
                     /db_xref="GI:54401411"
                     /translation="MNNLLRKIYSVLKYCTLKKFDTSTDYGRQQERYRIISLSIIANL
                     ISKFLSLLTLVLTVSLTLPYLGPARFGVWMTIASLSAALTFLDLGIGNALTNRIAHAC
                     ANGERTEVIKKTSGGIALLLLVSICITCVLYIICSVIDWRIFIKGINDNILGEVQRAI
                     KLFVVIFGISILSNGIQKIYMGMQKAYITYFFNSLLSLISIFCLIYYSSLHATIPVLI
                     LISSGIPLFTGMLLMILLIKDGFVSLNEIFHNMRNEAPHLISTGFFFFILQIGTLATW
                     SGDNFIISSTLGVAFVAILSLTQRLFQISTVPLTIYNTPLWVAYADAHARKDTHFIKK
                     TLLASFKIVGFSSLIMATFLILFGRDIISVWTGDKIDIPQSFIIAYAIWSIIDAVSNT
                     LASFLNGLGIVKQQMFAVVSLVIISIPAKYFIVSHFGLTAMLSCFIAIYIVNYIVWYT
                     ISFKKCIYSQLSI"
     gene            5188. .6285
                     /gene="per"
     CDS             5188. .6285
                     /gene="per"
                     /codon_start=1
                     /transl_table=11
                     /product="perosamine synthetase"
                     /protein_id="AAV34497.1"
                     /db_xref="GI:54401412"
                     /translation="MKYIPVYEPSLTGKEKEYVNDCLDSTWISSKGSYIQRFETEFSK
                     HISAQYATTVSNGTVAIHLALIALGISEGDEVIVPTLTYIASVNAINYTGAIPVFVDS
                     DAETWQVSIEDIRNKITNKTKAIMCVHLYGQPCDMDAILEIANEFNLFVIEDCAEAFG
                     SKYKDKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYERCVHYKGQGLAAHRQY
                     WHDVIGYNYRMTNICAAIGLAQLEQAEGFINRKRNIAEIYKKELFNSPVKVHQESDNT
                     FHTYWMVSILTRTAEEREELRNYLSENFIETRPLFYPVHTMPMYSEKYQKHPVAEDLG
                     WRGMNLPSYPGMTDEQVIYICNKIKCFYNEK"
     gene            6310. .7524
                     /gene="wbdP"
     CDS             6310. .7524
                     /gene="wbdP"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAV34498.2"
                     /db_xref="GI:55228557"
                     /translation="MKITINTDGFNEWGGGVDFIKYLLLILSTKDDVFTDIIVPKDDF
                     ASLFRRNVFPIKNAIKSLGKGKFPRWVKKDGFDEGYYHNAFSGLVDKSKLTFINSNSN
                     SYFDYHAKSDNDIIFPCMRVPTNSVCDSPWIGYIYDFQHCYYPSFFSAREIKKRELYF
                     REMLSSANNVIVNANSVIADAKNFIGDFSADLHALPFSPCPQQIWLDDNSNLVDKYAI
                     SKDYFIVCNQFWKHKNHVVVFKAFKEYIKVNPEAYLVCTGGIQDYRFPEYFGELKALI
                     QKLGISSKIKILGHIPKLEQVELIKKSIAVIQPTLFEGGPGGGATYDAVALGKKVILS
                     DIDVNREVNCGDVIFFNPQDEYSLAKALFKAEEEKSVKENGDLMKLGLKRREKCADFI
                     MSVIRKEIESRN"
     gene            7546. .8664
                     /gene="gmd"
     CDS             7546. .8664
                     /gene="gmd"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-mannose dehydratase"
                     /protein_id="AAV34499.1"
                     /db_xref="GI:54401414"
                     /translation="MSKVALITGVTGQDGSYLAEFLLEKGYEVHGIKRRASSFNTGRV
```

```
                     DHIYQDPHLGNPKFNLHYGDLTDTSNLTRILKEVQPDEVYNLGAMSHVAVSFESPEYT
                     ADVDAMGTLRLLEAIRFLGLEKKTRFYQASTSELYGLVREIPQKETTPFYPRSPYAVA
                     KLYAYWITVNYRESYGMYACNGILFNHESPRRGETFVTRKITRAMANIAQGLESCLYL
                     GNMDSLRDWGHAKDYVRMQWMMLQQEQPEDFVIATGVQYSVRDFVEMSAAQLGIKLSF
                     EGDGVNEKGIVASVTGHDAPGVKVGDVIVEVDPRYFRPAEVETLLGDPSKAHEKLGWK
                     PEITLQEMVAEMVASDLEAAKKYSLLKESGFDVNLSLE"
     gene            8667. .9632
                     /gene="fcl"
     CDS             8667. .9632
                     /gene="fcl"
                     /codon_start=1
                     /transl_table=11
                     /product="fucose synthetase"
                     /protein_id="AAV34500.1"
                     /db_xref="GI:54401415"
                     /translation="MKKQRIFVAGHRGMVGTAIVRQLSLRDNVELVLRTRDELNLLDS
                     SAVQAFFATERIDQVYLAAAKVGGIVANNTYPADFIYENMMIESNIIHAAHLHNVNKL
                     LFLGSSCIYPKQATQPIAESELLQGTLEPTNEPYAIAKIAGIKLCESYNRQYGRDYRS
                     VMPTNLYGPHDNFHPSNSHVIPALLRRFHEAREQNAPDVVVWGSGTPMREFLHVDDMA
                     AASIHVMELDSEIWQEYTQPMLSHINVGTGVDCTIRELAQTIAQVVGYKGKVVFDSSK
                     PDGTPRKLLDVTRLHSLGWRHSVSLEYGLESTYQWFLENQHWCRG"
     gene            9635. .10135
                     /gene="gmm"
     CDS             9635. .10135
                     /gene="gmm"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-mannose mannosylhydrolase"
                     /protein_id="AAV34501.1"
                     /db_xref="GI:54401416"
                     /translation="MFLHHDEFATVVRSTPLVSIDLIVENADGEFLLGKRLNRPAQGY
                     WFVPGGRVQKDERIENAFERLTLAELGLRLPLSVAEFYGVWQHFYDDNFSGPDFSTHY
                     IVLAFHIRGQESCLCLPDEQHKIWRWFKSEELEEEKDVHVNSRAYFHIENSNVIGLDN
                     KGLNHV"
     gene            10128. .11576
                     /gene="manC"
     CDS             10128. .11576
                     /gene="manC"
                     /codon_start=1
                     /transl_table=11
                     /product="mannose-1-P guanosyltransferase"
                     /protein_id="AAV34502.1"
                     /db_xref="GI:54401417"

  Query Match             41.1%;  Score 321.6;  DB 1;  Length 14224;
  Best Local Similarity   64.5%;  Pred. No. 1.3e-40;
  Matches  480;  Conservative    0;  Mismatches  264;  Indels    0;  Gaps    0;

Qy          2 TGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCT 61
              || | |  ||| | |||||||| |||||||| || || ||||||||     || ||  |
Db       1105 TGGAAACTGAATCAGTTTCAATCATTATGCCAGTATATAATGGGGCATCTACAATTGTAC 1164

Qy         62 CATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACG 121
                ||| ||||||| ||  || ||||  | ||| | ||  |     || ||||| ||||| |
Db       1165 AGTCAATAGAATCCATATTAAATCAGACCTATAAGGACATAAGGTTATATATTATTGATG 1224

Qy        122 ATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAA 181
               |||| || || ||||| || |    |||| |  | | |  ||| |  |  ||| | |
Db       1225 ATTGCAGTACTGATGAAACGTCAATATTACTACAAAATTTATATATTAGTAATGATAGGG 1284

Qy        182 TAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAG 241
              |    |  |  | || |  | |||   ||| || || |||||| |||||   ||||| |
Db       1285 TTGAGTTGATAAGAAATATCAAAAACAAAGGCGTCGCGGAAAGTAGAAATATAGGAATCG 1344

Qy        242 AAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAA 301
              ||  | | |  || |||||||  |||| |||||||| |||||||| |||   || |  |
Db       1345 AACAGTCTAGAGGCAAATATATAGCTTTCTGTGATGCTGATGATTTATGGGGTGAAACGA 1404

Qy        302 AATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATT 361
              |||||||    ||  | ||  |  |  |    |||   || |||||||| ||| | ||||
Db       1405 AATTAGAAAAACAGCTAGATATACTCGAATCAGAAAAAGTTGATGTGGTTTGTACAAATT 1464

Qy        362 ATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAA 421
              |||||  | ||||||    |    |||||| | |||   ||   | ||||     | || |
Db       1465 ATTATCTCATAGATTTGGAGGATAATATTATAGGCTCTGTATGTTCTCCAGGAATTATTA 1524

Qy        422 ATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCA 481
              |||||  | ||||||    |||||  | || |||||||| || || || |||||||  |
Db       1525 CTTATAAGACAATGCTAGGTAAAAATCATATTGGGAATTTAACGGGTATATATAATGTAA 1584

Qy        482 ACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGG 541
              | ||| | || ||  |||||||||| || ||||| || || |||||||| |||||| | |
Db       1585 AAAAAATAGGAAAAGTTTATCAAAAGAAAATTGGCCATGAAGATTATTTAATGTGGTTAG 1644

Qy        542 AAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGC 601
              ||||  | |  |||||  |    || || ||||| ||||| ||||| ||  | || | |
Db       1645 AAATCCTAAGAAAAACTGAACAAGCAATATGTATCCAAGAAAATCTTGCAAAATATAGGG 1704

Qy        602 GTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATAT 661
              |||||  || ||| |||| |  |||||| | | ||| ||| ||||||| |||| || |
Db       1705 TGTCAAAAAAATCATTATCAGCAAATAAATTAATAGCAGCACAATGGACGTGGAACATTT 1764

Qy        662 ATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCTT 721
              | ||  |||| || |  |  |  |  |  | |||||||||  ||||    |||||||
Db       1765 ACAGGAAACAACTAAAAATCCCAATAATACCAGCATTATATTTCTTTTCCCAATATGCTG 1824

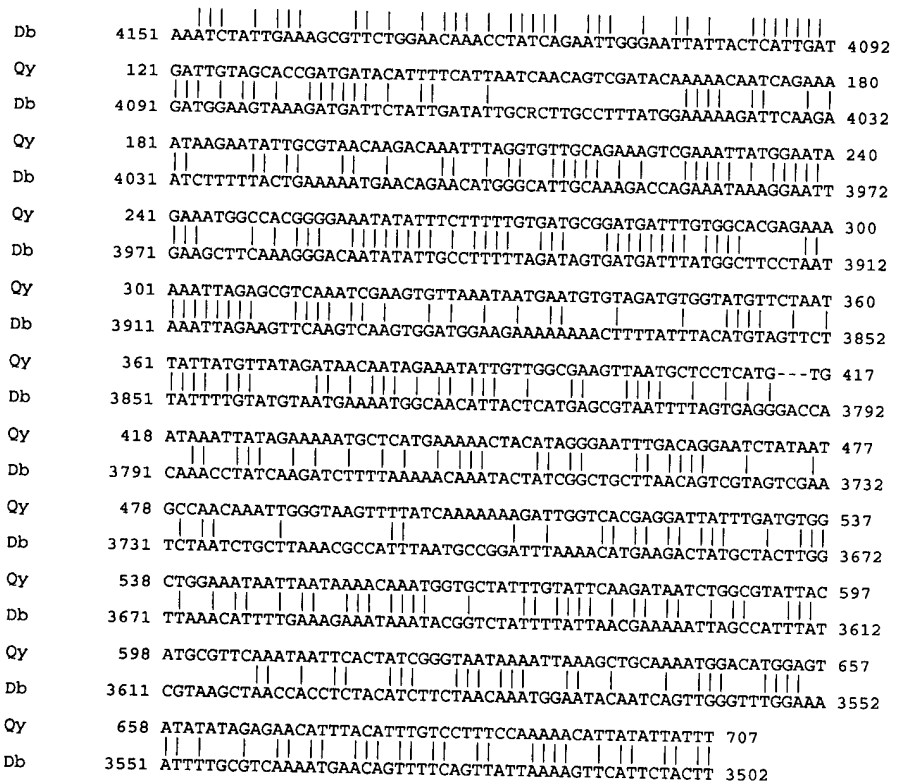
Qy        722 CAAATGGAGTCATGAAAAAAATAA 745
                | ||  |  | ||||||||||
Db       1825 TCCAAGGTTTATTAAAAAAAATAA 1848


RESULT 12
BD193531/c
LOCUS       BD193531               4915 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Enterococcus faecialis polynucleotides and polypeptides.
ACCESSION   BD193531
VERSION     BD193531.1  GI:33003270
KEYWORDS    JP 2002529046-A/125.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 4915)
  AUTHORS   Kunsch,C.A., Dillon,P.J. and Barash,S.C.
  TITLE     Enterococcus faecialis polynucleotides and polypeptides
  JOURNAL   Patent: JP 2002529046-A 125 03-SEP-2002;
            HUMAN GENOME SCIENCES INC
COMMENT     OS   Unidentified
            PN   JP 2002529046-A/125
            PD   03-SEP-2002
            PF   04-MAY-1998 JP 1998548302
            PR   06-MAY-1997 US    60/044031,16-MAY-1997 US    60/046655 PR
            14-NOV-1997 US    60/066009
            PI   CHARLES A KUNSCH,PATRICK J DILLON,STEVEN C BARASH PC
            C12N15/31,C07K14/315,C07K16/12,C12Q1/68
            CC   Strandedness: Double;
            CC   Topology: Linear;
            CC   Enterococcus faecialis polynucleotides and polypeptides FH
            Key             Location/Qualifiers
            FT   source          1. .4915
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .4915
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             17.9%;  Score 140.4;  DB 6;  Length 4915;
  Best Local Similarity   50.8%;  Pred. No. 1.5e-12;
  Matches  361;  Conservative    0;  Mismatches  346;  Indels    3;  Gaps    1;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||| | | || |   |||||||| ||||||||| | |||||||  | ||||  | ||
Db       4211 ATGGAAACAGCACTTGTTTCAATTATTATGCCCATGTACAATGCTGGCAAATTTTTATCA 4152

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
```



RESULT 13
AE016953/c
LOCUS AE016953 302241 bp DNA linear BCT 28-MAR-2003
DEFINITION Enterococcus faecalis V583, section 7 of 11 of the complete genome.
ACCESSION AE016953 AE016830
VERSION AE016953.1 GI:29343879
KEYWORDS .
SOURCE Enterococcus faecalis V583
ORGANISM Enterococcus faecalis V583
Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
REFERENCE 1 (bases 1 to 302241)
AUTHORS Paulsen,I., Banerjei,L., Myers,G.S.A., Nelson,K.E., Seshadri,R., Read,T.D., Fouts,D.E., Eisen,J.A., Gill,S.R., Heidelberg,J.F., Tettelin,H., Dodson,R.J., Umayam,L., Brinkac,L., Beanan,M., Daugherty,S., DeBoy,R.T., Durkin,S., Kolonay,J., Madupu,R., Nelson,W., Vamathevan,J., Tran,B., Upton,J., Hansen,T., Shetty,J., Khouri,H., Utterback,T., Radune,D., Ketchum,K.A., Dougherty,B.A. and Fraser,C.M.
TITLE Role of Mobile DNA in the Evolution of Vancomycin-Resistant Enterococcus faecalis
JOURNAL Science 299 (5615), 2071-2074 (2003)
PUBMED 12663927
REFERENCE 2 (bases 1 to 302241)
AUTHORS Paulsen,I., Banerjei,L., Myers,G.S.A., Nelson,K.E., Seshadri,R., Read,T.D., Fouts,D.E., Eisen,J.A., Gill,S.R., Heidelberg,J.F., Tettelin,H., Dodson,R.J., Umayam,L., Brinkac,L., Beanan,M., Daugherty,S., DeBoy,R.T., Durkin,S., Kolonay,J., Madupu,R., Nelson,W., Vamathevan,J., Tran,B., Upton,J., Hansen,T., Shetty,J., Khouri,H., Utterback,T., Radune,D., Ketchum,K.A., Dougherty,B.A. and Fraser,C.M.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES Location/Qualifiers
source 1. .302241
/organism="Enterococcus faecalis V583"
/mol_type="genomic DNA"
/strain="V583"
/db_xref="taxon:226185"
gene complement(260. .427)
/locus_tag="EF1887"
CDS complement(260. .427)
/locus_tag="EF1887"
/note="The gene assignment is based in part on a multiple alignment.; similar to GB:M99437, GB:U50549, PID:1293643, and PID:189264; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAO81640.1"
/db_xref="GI:29343880"
/translation="MTEKEKVEEIMEKYNRNFSTLQKNASAKELKTVFKFVADESNRKQRELIGLDKEK"
gene complement(420. .551)
/locus_tag="EF1888"
CDS complement(420. .551)
/locus_tag="EF1888"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO81641.1"
/db_xref="GI:29343881"
/translation="MTINILVKDFQNTISDFLMDIDNVELTLPISYTKEILEEELHD"
gene complement(655. .1095)
/locus_tag="EF1889"
CDS complement(655. .1095)
/locus_tag="EF1889"
/note="similar to GB:M21054, SP:P04054, PID:190013, and PID:387025; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAO81642.1"
/db_xref="GI:29343882"
/translation="MTEDDSLQYAKENKKEIISSVIDGKEKEEEKTAIFMAGSPGAGKTEAAQTLTVLNSNLCVIDADKFRVLFPGYVGNNSDEFQRGSSLLVDAALDLVLKKGYSFILDATFATSKVKQNIERALKKNYNVLVYYVYQDPFIAWDFTKK"
gene complement(1182. .1538)
/locus_tag="EF1890"
CDS complement(1182. .1538)
/locus_tag="EF1890"
/note="similar to GB:D14582, SP:P32856, and PID:303605; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAO81643.1"
/db_xref="GI:29343883"
/translation="MLTTKESAVILNKLKQIVMLGRQSGFFLILACQRPDAKYLGDGIRDQFNFRVALGRMSELGYSMMFGEVDKNFFMKRIKGRGYVDTGGSVISEFYTPLVPKGYDFLESIKQVAQSKEK"
gene complement(1640. .3456)
/locus_tag="EF1891"
/note="This region contains an authentic frame shift and is not the result of a sequencing artifact; group II intron, maturase, putative, authentic frameshift; similar to GP:5360569; identified by sequence similarity; putative"
gene complement(4107. .5123)

```
                /locus_tag="EF1892"
     CDS        complement(4107. .5123)
                /locus_tag="EF1892"
                /note="similar to GB:D14582, SP:P32856,  and PID:303605;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="FtsK/SpoIIIE family protein"
                /protein_id="AAO81644.1"
                /db_xref="GI:29343884"
                /translation="MYKEHRIRARDQHLVYHFILGWLIALLISWMGVFYFQEFRQFDI
                SRVSLSTIETVWSMKELICLLGSLGFSGAMLLLYIHFFPDHWRSLWHRQKLARMILEN
                HWYEVKQTQSEGFFKDLNSSRTRETISYFPKIYYRMKEGLLSIRVQISLGKYQEQLLK
                LEKKLESGLYCELVEKELKDSYVEYTLLYDMIANRIGIDEVVAENGTLRLMKNQVWAY
                DSLPHMLIAGGTGGGKTYFLLTIIEALLKSDAELFILDPKNADLADLGTVMPHVYSQK
                EEISACVEDFYERMIARSKAMKEMPNYKPGENYAYLGLPPNFLIFDEYVAYMGANRFP
                TSIE"
     gene       complement(5193. .5546)
                /locus_tag="EF1893"
     CDS        complement(5193. .5546)
                /locus_tag="EF1893"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAO81645.1"
                /db_xref="GI:29343885"
                /translation="MKIMEINVEHILDCLDQYGKGELTEEQLVKALTYDEKMFLIMHQ
                GLLGENSDSEDDFDVQDWLEEEESFFMVVEVNENLCRQAESVLEEIGVEMPDAIEGFL
                NQLVEKKQFPVVINS"
     gene       complement(5547. .5921)
                /locus_tag="EF1894"
     CDS        complement(5547. .5921)
                /locus_tag="EF1894"
                /note="similar to GB:L04953, SP:Q02410,  and PID:340409;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAO81646.1"
                /db_xref="GI:29343886"
                /translation="MRLTEGIVVDSVLTFGKLRFSALRREVRKQNEDGTVSNEVKERT
                YNLKSSAQGRMIQVSIPANVPLREFAYDAEVELVNPIVDTVANYVFREGTTVNWFIKA
                DDLVLKRQPNQGNSTNQIEGKK"
     gene       complement(5934. .6248)
                /locus_tag="EF1895"
     CDS        complement(5934. .6248)
                /locus_tag="EF1895"
                /note="similar to GB:X05615, GB:X06065, GB:X02154,
                SP:P01266, PID:1335349, PID:1359884, PID:2204216,  and
                PID:37174; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAO81647.1"
                /db_xref="GI:29343887"
                /translation="MELKFVVPDMAETFGKLSYAGEGEILTEGYGRNTTVIGRSYHLY
                SSKQRADDIEVVVAAEAGEKDFDQDQPLKAVNPHLVAKGYEIENRGFTDYVLYVDDLV
                KE"
     gene       complement(6361. .9588)
                /locus_tag="EF1896"
     CDS        complement(6361. .9588)
                /locus_tag="EF1896"
                /note="identified by match to TIGR protein family HMM
                TIGR01167"
                /codon_start=1
                /transl_table=11
                /product="cell wall surface anchor family protein"
                /protein_id="AAO81648.1"
                /db_xref="GI:29343888"
                /translation="MIKPIFKKWRLFAALALLGQTIVGAIDPMIVFADEIAHPQTVTV
                ELDLAHQYVVEGTFSDGRPMSEVTVPHYAVYNGVKQDVFCIEPGVPIDNEFTPGYEKN
                PLPDMPEKAKLVSVLWKKAGTDVDTHIVAQKMIWQEVNGYTLHSIKRLNGSAVNIAAI
                EAKINQAIADYQKKPSFHNSTAKIVLGQSTTMTDTNNLNLSEFDEVVENTANIDYRVN
                GNQLIITPNASSKESGVLTLKKSAGTGTPVAYKKVGQQTLMAGAIDKPNTYTVKIDVE
                TEGSLKIKKVDKESGAIVPGTVFHLDFGKTLPAKDVTTDKEGIATLDGIPHGTKVTIT
                EKSVPAPYTIDTIPMTATIKAGETIFMTSKNTREKGQIILDKTGIETGSDLWNDNYSL
                AGNTFAIRKDSPTGEIVQEMTTDEKGHAETPKEIANALELGTYYVTETKASNGFVNTF
                KAVKVELKYANQTVALVTSNIKGQNQEITGETTLIKEDKDTGDKTQGRAVFEGTEYTL
                FTAKDGKAVKWSEAFKPELVKGTKASDETVTLALDEKNQAAVKHLAINEYYWQETKAP
                EGYTLDETKYPVSIKKVDDNEKNAVITRDVTAKEQIIRFGFDFFKFAGSAAGTAETGF
                NDLTFKVSPLEGTNEITGAEDEAITAYNEQLGFDGYGKFENLPYGDYLLEEVEAPEGF
                QKITPLEIRSTFKENKEDFVKSEYVFTITEQGQKQPIKTVTVPYEKLTNKAFSVSLNR
                LMLYDLPEEEDSLTSLATWKDGNKKLNTLDFTELVDKLRYNLHEIKEDWYVVAQAIDV
                EATKAAQEKDEKAKPVVIAETTTTLANKEKTGTWKIRHKLTTEQVLDKTIVLFNYVYE
                NKEAFEAGDEPVAKDASLNNQAQTVNCTVERHVSIQTKAHLEDGSQTFTHGDVVDMFD
                DVSITHDVLDGSKEAFETILYALLPDGTNKEIWKSGKIDYEVNDKEFTKTMLAEKVDT
                GKFPESTKFTFAEINYDKDENINGKHNEDLKEKNQTLTPKKVPTTPRTPKQPETPTVP
                SGYQESSPTVKTFPQTGEKNSNVLLFIGFTLIFATAGYYFWNRRN"
     gene       complement(9668. .10330)
                /locus_tag="EF1897"
     CDS        complement(9668. .10330)
                /locus_tag="EF1897"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAO81649.1"
                /db_xref="GI:29343889"
                /translation="MKQDDGRIVWKNLSDLKLILLINQFIEKHEIKSSRQYHRKLLEN
```

```
  Query Match             17.9%;  Score 140.4;  DB 1;  Length 302241;
  Best Local Similarity   50.8%;  Pred. No. 3e-13;
  Matches  361;  Conservative    0;  Mismatches  346;  Indels    3;  Gaps    1;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||| | | || |   |||||||| ||||||||| | |||||||  | ||||  | ||
Db     256097 ATGGAAACAGCACTTGTTTCAATTATTATGCCCATGTACAATGCTGGCAAATTTTTATCA 256038

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
                |||  | |||    || | | ||| | |||||  |||  |  || |   |||||||
Db     256037 AAATCTATTGAAAGCGTTCTGGAACAAACCTATCAGAATTGGGAATTATTACTCATTGAT 255978

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||| | || |  |||||| |  ||    |                ||||  ||  | |
Db     255977 GATGGAAGTAAAGATGATTCTATTGATATTGCACTTGCCTTTATGGAAAAAGATTCAAGA 255918

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||      || ||   ||  |    ||  | ||  ||||| | |   ||||| | |||||
Db     255917 ATCTTTTTACTGAAAAATGAACAGAACATGGGCATTGCAAAGACCAGAAATAAAGGAATT 255858

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              |||    | | |||  |||||||| | ||||  |||   |||||||| ||||    ||
Db     255857 GAAGCTTCAAAGGGACAATATATTGCCTTTTTAGATAGTGATGATTTATGGCTTCCTAAT 255798

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||   |||| || |   | |  |  ||    |      |   ||||  | |
Db     255797 AAATTAGAAGTTCAAGTCAAGTGGATGGAAGAAAAAAAACTTTTATTTACATGTAGTTCT 255738

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATG---TG 417
              |||| |||      || | ||| | || |||  |   ||   ||||  |  | | |
Db     255737 TATTTTGTATGTAATGAAAATGGCAACATTACTCATGAGCGTAATTTTAGTGAGGGACCA 255678

Qy        418 ATAAATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAAT 477
                ||  ||| | |  |  | |  | |||    || ||   || ||||   |     |
Db     255677 CAAACCTATCAAGATCTTTTAAAAACAAATACTATCGGCTGCTTAACAGTCGTAGTCGAA 255618

Qy        478 GCCAACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGG 537
              | ||     |          ||        |  |    || || || |||    |  |||
Db     255617 TCTAATCTGCTTAAACGCCATTTAATGCCGGATTTAAAACATGAAGACTATGCTACTTGG 255558

Qy        538 CTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTAC 597
              |  | ||  | ||  ||| ||||    |    || |||| | || ||  | ||   |||
Db     255557 TTAAACATTTTGAAAGAAATAAATACGGTCTATTTTATTAACGAAAAATTAGCCATTTAT 255498

Qy        598 ATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGT 657
```

```
Db      255497 CGTAAGCTAACCACCTCTACATCTTCTAACAAATGGAATACAATCAGTTGGGTTTGGAAA 255438

Qy         658 ATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTT 707

Db      255437 ATTTTGCGTCAAAATGAACAGTTTTCAGTTATTAAAAGTTCATTCTACTT 255388
```

RESULT 14
AE017355_15
WPCOMMENT
Sequence split into 53 fragments LOCUS AE017355 Accession AE017355

Fragment Name	Begin	End
AE017355_00	1	110000
AE017355_01	100001	210000
AE017355_02	200001	310000
AE017355_03	300001	410000
AE017355_04	400001	510000
AE017355_05	500001	610000
AE017355_06	600001	710000
AE017355_07	700001	810000
AE017355_08	800001	910000
AE017355_09	900001	1010000
AE017355_10	1000001	1110000
AE017355_11	1100001	1210000
AE017355_12	1200001	1310000
AE017355_13	1300001	1410000
AE017355_14	1400001	1510000
AE017355_15	1500001	1610000
AE017355_16	1600001	1710000
AE017355_17	1700001	1810000
AE017355_18	1800001	1910000
AE017355_19	1900001	2010000
AE017355_20	2000001	2110000
AE017355_21	2100001	2210000
AE017355_22	2200001	2310000
AE017355_23	2300001	2410000
AE017355_24	2400001	2510000
AE017355_25	2500001	2610000
AE017355_26	2600001	2710000
AE017355_27	2700001	2810000
AE017355_28	2800001	2910000
AE017355_29	2900001	3010000
AE017355_30	3000001	3110000
AE017355_31	3100001	3210000
AE017355_32	3200001	3310000
AE017355_33	3300001	3410000
AE017355_34	3400001	3510000
AE017355_35	3500001	3610000
AE017355_36	3600001	3710000
AE017355_37	3700001	3810000
AE017355_38	3800001	3910000
AE017355_39	3900001	4010000
AE017355_40	4000001	4110000
AE017355_41	4100001	4210000
AE017355_42	4200001	4310000
AE017355_43	4300001	4410000
AE017355_44	4400001	4510000
AE017355_45	4500001	4610000
AE017355_46	4600001	4710000
AE017355_47	4700001	4810000
AE017355_48	4800001	4910000
AE017355_49	4900001	5010000
AE017355_50	5000001	5110000
AE017355_51	5100001	5210000
AE017355_52	5200001	5237682

Continuation (16 of 53) of AE017355 from base 1500001 (AE017355 Bacillus thuringiensis s

Query Match 17.0%; Score 133.4; DB 1; Length 110000;
Best Local Similarity 49.7%; Pred. No. 5.3e-12;
Matches 369; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

```
Qy          16 GTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCA 75

Db       45639 GTTTCTGTAATAACACCTTCTTATAATAGTATACGATTTATAGGTGAGACGATTGTATCT 45698

Qy          76 ATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGAT 135

Db       45699 GTACAAAATCAATCATATGAAAATTGGGAGATGATTATCGTTGATGACGCTTCAACTGAC 45758

Qy         136 GATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGT 195

Db       45759 GAATCTGTTACAAAAATTAAAGAGATAATAGAAGGAGACTCGCGTATTAGGTTAGTATCA 45818

Qy         196 AACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGG 255

Db       45819 TTAAAAGAAAATATTGGTGCTGCTAAGGCTCGGAATATAGCAATTCAAGAGGCGAGAGGA 45878

Qy         256 AAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAA 315

Db       45879 AGGTATATTGCCTTTTTAGATAGTGATGATATATGGTTACCGCATAAATTGAAGACACAA 45938

Qy         316 ATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGAT 375

Db       45939 TTGTTATTTATGGAAGAAATGAGTGTGGCCTTTTCATATGCATCGTATAGTTTAATCGAT 45998

Qy         376 AACAATAGAAATATTGT---TGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAA 432

Db       45999 GAAAACGGTAATGGACTAAATCGAGAAGTGAATGTACCGAAATCTGTTGACTATCATTGT 46058

Qy         433 ATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTGGGT 492

Db       46059 TTGGCAGGAAATACAATTATCGGATGTTTAACAGTAATAATTGATCGTGAAAAAATTCCG 46118

Qy         493 AAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGGAAATAATTAAT 552

Db       46119 CATATTGAAATGCCTAGTGTACAGCCGGAAGATACGGCGTTATGGCTGAAATTATTACAT 46178

Qy         553 AAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCAAATAAT 612

Db       46179 GAAGGGCATGAAGCGAAAGGGATACAGCAAGTATTAGCGAAGTATCGAATTGTAGCAAAT 46238

Qy         613 TCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATATATAGAGAACAT 672

Db       46239 TCGGTTTCCAGAAATAAAATTAAAGCAGCTTTTCGGTATTGGAAATTATTAAGAGACCAA 46298

Qy         673 TTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCTTCAAATGGAGTC 732

Db       46299 AAATGTCTTAATGCAGTGCAAATCTTTTACTATTTTAGTAAGTATGCTTATCATGCCTAT 46358

Qy         733 ATGAAAAAAATAACACATTCACT 755

Db       46359 AGAAAAAATAAAATCAATGTAGT 46381
```

RESULT 15
AE017225_15
WPCOMMENT
Sequence split into 53 fragments LOCUS AE017225 Accession AE017225

Fragment Name	Begin	End
AE017225_00	1	110000
AE017225_01	100001	210000
AE017225_02	200001	310000
AE017225_03	300001	410000
AE017225_04	400001	510000
AE017225_05	500001	610000
AE017225_06	600001	710000
AE017225_07	700001	810000
AE017225_08	800001	910000
AE017225_09	900001	1010000
AE017225_10	1000001	1110000
AE017225_11	1100001	1210000
AE017225_12	1200001	1310000
AE017225_13	1300001	1410000
AE017225_14	1400001	1510000
AE017225_15	1500001	1610000

AE017225_16	1600001	1710000
AE017225_17	1700001	1810000
AE017225_18	1800001	1910000
AE017225_19	1900001	2010000
AE017225_20	2000001	2110000
AE017225_21	2100001	2210000
AE017225_22	2200001	2310000
AE017225_23	2300001	2410000
AE017225_24	2400001	2510000
AE017225_25	2500001	2610000
AE017225_26	2600001	2710000
AE017225_27	2700001	2810000
AE017225_28	2800001	2910000
AE017225_29	2900001	3010000
AE017225_30	3000001	3110000
AE017225_31	3100001	3210000
AE017225_32	3200001	3310000
AE017225_33	3300001	3410000
AE017225_34	3400001	3510000
AE017225_35	3500001	3610000
AE017225_36	3600001	3710000
AE017225_37	3700001	3810000
AE017225_38	3800001	3910000
AE017225_39	3900001	4010000
AE017225_40	4000001	4110000
AE017225_41	4100001	4210000
AE017225_42	4200001	4310000
AE017225_43	4300001	4410000
AE017225_44	4400001	4510000
AE017225_45	4500001	4610000
AE017225_46	4600001	4710000
AE017225_47	4700001	4810000
AE017225_48	4800001	4910000
AE017225_49	4900001	5010000
AE017225_50	5000001	5110000
AE017225_51	5100001	5210000
AE017225_52	5200001	5228663

Continuation (16 of 53) of AE017225 from base 1500001 (AE017225 Bacillus anthracis str.

Query Match 16.6%; Score 129.8; DB 1; Length 110000;
Best Local Similarity 50.2%; Pred. No. 1.9e-11;
Matches 373; Conservative 0; Mismatches 367; Indels 3; Gaps 2;

```
Qy         16 GTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCA 75
              |||||  |||| |  ||   ||| ||| |       |  ||||        |  | | |||
Db      19452 GTTTCTGTAATAACACCTTCTTATAATAGTATACGATTTATAGGTGAGACGATTGTATCT 19511

Qy         76 ATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGAT 135
                |  | ||||||| ||| || |||  |   ||  |||| |||| ||   |   || ||
Db      19512 GTACAAAATCAATCATATGAAAATTGGGAGATGATTATCGTTGATGACGCTTCAACTGAC 19571

Qy        136 GATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGT 195
              ||  |  || ||  ||| ||     |    ||    |   |   || ||  || |
Db      19572 GAATCTGTTACAAAAATTAAAGAGATAATAGAAGGAGACTCGCGTATTAGGTTAGTATCA 19631

Qy        196 AACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGG 255
               ||   |||| | |||| |||  |   ||| |||   | |||  ||  ||| |  ||
Db      19632 TTAAAAGAAAATATTGGTGCTGCTAAGGCTCGGAATATAGCAATTCAAGAGGCGAGAGGA 19691

Qy        256 AAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAA 315
              |  |||||| | ||||  |||   ||||||| | |||      | | |||||  ||   |||
Db      19692 AGGTATATTGCCTTTTTAGATAGTGATGATATATGGTTACCGCATAAATTGAAGACACAA 19751

Qy        316 ATCGAAGTGTT--AAATAATGAATGTGTAGATGTGGTATG-TTCTAATTATTATGTTATA 372
               |   | |  |  ||  ||||||||||    | |  ||||  ||| ||  ||   ||
Db      19752 TTGTTATTTATGGAAGAAATGAATGTGTCCTTTTCATATGCATCTTATAGTTTAATTGAT 19811

Qy        373 GATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAA 432
              || |||  ||   |  |   | |  |||| ||||  ||  |     |  | |||
Db      19812 GAAAACGGTAATGAACTAAATCGAAAAGTGAATGTACCGAAATCTGTTGACTATCATTGT 19871

Qy        433 ATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTGGGT 492
               ||     ||| |     || ||   ||| ||||  ||   | ||    | | | |
```

```
Db      19872 TTGGCAGGGAATACAATTATCGGATGTTTAACAGTGATAATTGATCGTGAAAGAATTCCG 19931

Qy        493 AAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGGAAATAATTAAT 552
               |  || |       |   |    |  || |||     | | |||||| || || |  ||
Db      19932 CATATTGAAATGCCTAGTGTACAGCCGGAAGATACGGCGTTATGGCTGAAATTATTACAT 19991

Qy        553 AAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCAAATAAT 612
               ||    |||  || |   | || ||  |      | ||  | ||      | |   |||
Db      19992 GAAGGGCATGAAGCGAAAGGGATACAGCAAGTATTAGCAAAGTATCGAATTGTAGCAAAT 20051

Qy        613 TCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATATATAGAGAACAT 672
              ||  | ||  | |||||||||||||| ||     ||   ||||   |||  ||||| ||
Db      20052 TCGGTTTCCAGAAATAAAATTAAAGCAGCTTTTCGGTATTGGAAATTATTAAGAGACCAA 20111

Qy        673 TTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCTTCAAATGGAGTC 732
                |  | |           |||  || || ||||||     |||||||   |||
Db      20112 AAATGTCTTAATGCAGTGCAAATCTTTTACTATTTTAGTAAGTATGCTTATCATGCCTAT 20171

Qy        733 ATGAAAAAAATAACACATTCACT 755
              |  ||||| | ||   ||  | |
Db      20172 AGAAAAAATAAAATCAATGTAGT 20194
```

Search completed: May 10, 2006, 20:33:38
Job time : 2849.13 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:42:47 ; Search time 365.515 Seconds
(without alignments)
14276.972 Million cell updates/sec

Title: US-09-423-093B-2_COPY_79_861
Perfect score: 783
Sequence: 1 atgaacaaagaaaccgtttc.....gaaaggagactaaaaagtga 783

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	783	100.0	14024	2	AAX06749	Aax06749 E. coli O
	2	783	100.0	14024	3	AAZ56386	Aaz56386 Escherich
c	3	781.4	99.8	14181	10	ADC01214	Adc01214 Enterohae
c	4	781.4	99.8	14187	9	ACD19108	Acd19108 E. coli 0
	5	544	69.5	18772	14	AEC31468	Aec31468 E. coli t
	6	181.8	23.2	747	8	ACA29283	Aca29283 Prokaryot
	7	162.8	20.8	774	8	ACA28974	Aca28974 Prokaryot
	8	162.8	20.8	12220	14	AEA88828	Aea88828 Escherich
c	9	140.4	17.9	4915	2	AAX13062	Aax13062 Enterococ
c	10	140.4	17.9	4915	6	ABS98857	Abs98857 Enterococ
	11	138.4	17.7	30248	13	ADV98191	Adv98191 Shigella
	12	126.6	16.2	10812	11	ADN10409	Adn10409 O-antigen
	13	116.8	14.9	753	8	ACA34280	Aca34280 Prokaryot
	14	116.8	14.9	110000	2	AAT42063_09	Continuation (10 o
	15	108.8	13.9	9227	11	ADN02611	Adn02611 S. dysent
	16	100	12.8	891	8	ABT13680	Abt13680 Campyloba
c	17	98.4	12.6	894	13	ADU77565	Adu77565 Beta-1,3-
c	18	98.4	12.6	12219	13	ADU77595	Adu77595 Campyloba
	19	96.4	12.3	13074	14	AEC21818	Aec21818 E. coli O
	20	94.2	12.0	48551	6	AAS20800	Aas20800 Clostridi
	21	93.8	12.0	3732	3	AAZ60932	Aaz60932 Nucleotid
	22	93.2	11.9	906	3	AAA53723	Aaa53723 Campyloba
	23	93.2	11.9	906	8	ABT13678	Abt13678 Campyloba
c	24	93.2	11.9	11474	3	AAA53720	Aaa53720 LPS core
c	25	93.2	11.9	11474	8	ABT13665	Abt13665 Campyloba
	26	88.6	11.3	110000	10	ADF77343_04	Continuation (5 of
	27	86.4	11.0	990	10	ACF71590	Acf71590 Photorhab
	28	86.4	11.0	110000	10	ACF67367_47	Continuation (48 o
c	29	86.4	11.0	110000	10	ACF65388_00	Acf65388 Photorhab
	30	86.2	11.0	807	9	ADB09559	Adb09559 Alloiococ
	31	86.2	11.0	807	9	ADB09561	Adb09561 Alloiococ
	32	86.2	11.0	807	9	ADB09557	Adb09557 Alloiococ
	33	86.2	11.0	110000	9	ADB12064_09	Continuation (10 o
c	34	85.2	10.9	110000	2	AAV21209_04	Continuation (5 of
	35	81.2	10.4	1002	8	ACA28115	Aca28115 Prokaryot
	36	80.8	10.3	19966	3	AAZ30355	Aaz30355 Nucleotid
	37	80.8	10.3	19966	3	AAZ45258	Aaz45258 DNA encod
	38	75.6	9.7	1011	8	ACA27781	Aca27781 Prokaryot
	39	74.8	9.6	2349	4	AAH43198	Aah43198 DNA encod
	40	74.8	9.6	6992	3	AAZ60930	Aaz60930 Nucleotid
	41	72.8	9.3	1002	10	ADC91842	Adc91842 E. faeciu
	42	69.8	8.9	15950	3	AAZ30357	Aaz30357 Nucleotid
	43	69.8	8.9	15950	3	AAZ45260	Aaz45260 DNA encod
	44	69.2	8.8	1774	12	ADG32133	Adg32133 DNA encod
	45	68.4	8.7	757	6	ABK74443	Abk74443 Bacillus

ALIGNMENTS

RESULT 1
AAX06749
ID AAX06749 standard; DNA; 14024 BP.
XX
AC AAX06749;
XX
DT 26-APR-1999 (first entry)
XX
DE E. coli O157 antigen gene cluster.
XX
KW O antigen; O157 antigen; diagnosis; wbdN gene; wzx gene; wzy gene;
KW wbdP gene; wbdO gene; wbdR gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Escherichia coli.
XX

FH	Key	Location/Qualifiers
FT	CDS	79. .861
FT		/*tag= a
FT		/product= "sugar_transferase"
FT		/note= "wbdN gene (ORF1), this region is specifically
FT		claimed in Claim 6; encodes AAW88312"
FT	CDS	858. .2042
FT		/*tag= b
FT		/product= "O_antigen-polymerase"
FT		/note= "wzy gene (ORF2), this region is specifically
FT		claimed in Claim 6; encodes AAW88313"
FT	CDS	2011. .2757
FT		/*tag= c
FT		/product= "sugar_transferase"
FT		/note= "wbdO gene (ORF3), this region is specifically
FT		claimed in Claim 6; encodes AAW88314"
FT	CDS	2744. .4135
FT		/*tag= d
FT		/product= "O_antigen_flippase"
FT		/note= "wzx gene (ORF4), this region is specifically
FT		claimed in Claim 6; encodes AAW88315"
FT	CDS	4132. .5232
FT		/*tag= e
FT		/product= "perosamine_synthetase"
FT		/note= "per gene (ORF5); encodes AAW88316"
FT	CDS	5257. .6471

```
FT                   /*tag=  f
FT                   /product= "sugar_transferase"
FT                   /note= "wbdP gene (ORF6), this region is specifically
FT                   claimed in Claim 6; encodes AAW88317"
FT   CDS             6491. .7609
FT                   /*tag=  g
FT                   /note= "ORF7; encodes AAW88318"
FT   CDS             7606. .8578
FT                   /*tag=  h
FT                   /note= "ORF8; encodes AAW88319"
FT   CDS             8580. .9089
FT                   /*tag=  i
FT                   /note= "ORF9; encodes AAW88320"
FT   CDS             9071. .10519
FT                   /*tag=  j
FT                   /note= "manC gene (ORF10); encodes AAW88321"
FT   CDS             10523. .11893
FT                   /*tag=  k
FT                   /note= "manD gene (ORF11); encodes AAW88322"
FT   repeat_region   12007. .13123
FT                   /tag= l
FT                   /note= "H-repeat, may play a role in gene cluster
FT                   assembly"
FT   CDS             13156. .13821
FT                   /*tag=  m
FT                   /product= "N-acetyl_transferase"
FT                   /note= "wbdR gene (ORF12), this region is specifically
FT                   claimed in Claim 6; encodes AAW88323"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88312, AAW88313, AAW88314, AAW88315, AAW88316, AAW88317,
DR   AAW88318, AAW88319, AAW88320, AAW88321, AAW88322, AAW88323.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 6; Fig 8; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Escherichia coli O157 O antigen. The gene cluster was
CC   obtained by PCR amplification of E. coli O157:H7 strain (C664-1992) DNA
CC   using primers (see AAX06752-63) based on O antigen gene sequences. 12
CC   Open reading frames within the gene cluster were identified on the basis
CC   of homology to known sequences. The genes encode proteins (see AAW88312-
CC   23) involved the synthesis of sugars present in the polysaccharide
CC   antigen, and in the transport or processing of polysaccharide or
CC   oligosaccharide units. The use of nucleic acid molecules derived from
CC   particular assembly and transport genes, particularly wbd (transferase),
CC   wzx (flippase) and wzy (polymerase) genes, within O antigen gene clusters
CC   improves the specificity of methods for the detection and identification
CC   of O antigens, e.g. in tests of food- or faecal- derived samples, or
CC   samples from patients. The O antigen is a major virulence factor of
CC   enteropathogenic E. coli strains that cause diarrhoea and haemorrhagic
CC   colitis
XX
SQ   Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 783;  DB 2;  Length 14024;
  Best Local Similarity 100.0%;  Pred. No. 8.9e-154;

  Matches  783;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 138

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 198

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 258

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 318

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 378

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 438

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 498

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 558

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 618

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 678

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 738

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 798

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 858

Qy        781 TGA 783
              |||
Db        859 TGA 861


RESULT 2
AAZ56386
ID   AAZ56386 standard; DNA; 14024 BP.
XX
AC   AAZ56386;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:56.
XX
KW   Flagellin; fliC; antigen; detection; ds.
XX
OS   Escherichia coli.
XX
```

PN WO9961458-A1.
XX
PD 02-DEC-1999.
XX
PF 21-MAY-1999; 99WO-AU000385.
XX
PR 21-MAY-1998; 98AU-00003634.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Reeves PR, Wang L;
XX
DR WPI; 2000-072598/06.
XX
PT Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, feces, etc.
XX
PS Claim 3; Page 229-233; 245pp; English.
XX
CC AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all or
CC part of an Escherichia coli flagellin protein except a protein expressed
CC by E. coli H1, H7, H12 or H48 type strains. The present invention also
CC describes a method of detecting the presence of E. coli of a particular H
CC serotype in a sample, comprising specifically hybridising a nucleic acid,
CC preferably at least a pair, derived from a flagellating gene, specific
CC for a particular flagellin gene associated with the H serotype, to any
CC E.coli in the sample which contain the gene, and detecting any hybridised
CC molecules, identifying the presence of that serotype in the sample. (I)
CC are useful for: (1) detecting the presence of E. coli of H serotype in a
CC sample by hybridising at least one or a pair of (I) to any E. coli in the
CC sample and detecting the hybridised nucleic acid molecules; and (2) for
CC detecting the presence of both O and H-serotypes of E. coli by
CC hybridising at least one or a pair of (I) to any E. coli present in the
CC sample and detecting the hybridised nucleic acid molecules. (I) is
CC particularly useful for detecting the combination of O and H antigen.
CC Hybridised (I) when using at least one (I) is detected by southern blot
CC analysis and, when using a pair of (I), is detected by polymerase chain
CC reaction (PCR). AAZ56399 to AAZ56420 represent primers used in the
CC exemplification of the present invention
XX
SQ Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 U; 0 Other;

Query Match 100.0%; Score 783; DB 3; Length 14024;
Best Local Similarity 100.0%; Pred. No. 8.9e-154;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 138

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 198

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 258

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 318

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 378

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 438

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 498

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 558

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 618

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 678

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 738

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 798

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 858

Qy        781 TGA 783
              |||
Db        859 TGA 861
```

RESULT 3
ADC01214/c
ID ADC01214 standard; DNA; 14181 BP.
XX
AC ADC01214;
XX
DT 04-DEC-2003 (first entry)
XX
DE Enterohaemorragic E. coli 0157:H7-specific nucleic acid SEQ ID NO: 1258.
XX
KW ds; gene; enterohaemorragic; anti-bacterial.
XX
OS Escherichia coli; 0157:H7.
XX
PN JP2002355074-A.
XX
PD 10-DEC-2002.
XX
PF 24-JAN-2002; 2002JP-00015959.
XX
PR 24-JAN-2001; 2001JP-00112010.
XX
PA (UYTS-) UNIV TSUKUBA.
XX
DR WPI; 2003-451640/43.
XX
PT Enterohemorragic Escherichia coli O157:H7-specific nucleic acid molecule
PT and a polypeptide and its use, a polypeptide, a vector and a host cell.
XX
PS Claim 2; SEQ ID NO 1258; 2067pp; Japanese.
XX
CC The invention relates to a novel enterohaemorragic Escherichia coli
CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
CC has anti-bacterial activity. The polypeptide can be used in detection
CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
CC genome of Enterohaemorragic E coli O157:H7 was determined. The present
CC sequence represents an E. coli 0157:H7-specific nucleic acid of the
CC invention.
XX
SQ Sequence 14181 BP; 4450 A; 2856 C; 2281 G; 4594 T; 0 U; 0 Other;

Query Match 99.8%; Score 781.4; DB 10; Length 14181;
Best Local Similarity 99.9%; Pred. No. 1.9e-153;

```
  Matches  782;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13917 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 13858

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13857 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 13798

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13797 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 13738

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13737 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 13678

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13677 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 13618

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13617 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 13558

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      13557 TATTATGTTATAGATAACAAAAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 13498

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13497 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 13438

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13437 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 13378

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13377 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 13318

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13317 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 13258

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13257 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 13198

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13197 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 13138

Qy        781 TGA 783
              |||
Db      13137 TGA 13135


RESULT 4
ACD19108/c
ID   ACD19108 standard; DNA; 14187 BP.
XX
AC   ACD19108;
XX
DT   27-OCT-2003  (revised)
DT   21-AUG-2003  (first entry)
XX
DE   E. coli 0157 unique DNA sequence OZID_121.
XX
KW   OZID; ds; acute haemorrhagic colitis; haemolytic uraemic syndrome;
KW   food poisoning.
XX
OS   Escherichia coli; strain 0157:H7.
XX
PN   US2003023075-A1.
XX
PD   30-JAN-2003.
XX
PF   01-APR-2002; 2002US-00114170.
XX
PR   04-DEC-1998;   98US-0110955P.
PR   03-DEC-1999;   99US-00453702.
XX
PA   (BLAT/) BLATTNER F R.
PA   (BURL/) BURLAND V D.
PA   (PERN/) PERNA N T.
PA   (PLUN/) PLUNKETT G.
PA   (WELC/) WELCH R.
XX
PI   Blattner FR,  Burland VD,  Perna NT,  Plunkett G,  Welch R;
XX
DR   WPI; 2003-479497/45.
XX
PT   New DNA sequences from Escherichia coli strain O157:H7, useful for
PT   detecting E. coli O157:H7 in a sample, or in designing diagnostic probes
PT   which can be used to distinguish strain O157:H7 from strain K12 using
PT   molecular techniques.
XX
PS   Claim 16; SEQ ID NO 121; 33pp; English.
XX
CC   The invention relates to an isolated DNA molecule comprising an E. coli
CC   strain 0157:H7 sequence selected from a clostridial cytotoxin-like gene,
CC   a urease gene cluster, a RTX toxin-like gene cluster, a locus of
CC   enterocyte effacement and 2 genes from its associated lymphocytic phage
CC   933W (a putative serine/threonine kinase and a tail fibre gene). E. coli
CC   0157:H7 can cause food poisoning, specifically acute haemorrhagic colitis
CC   (which can develop into haemolytic uraemic syndrome). Also included are
CC   an isolated DNA molecule comprising a nucleotide sequence identical to at
CC   least 25 contiguous nucleotides contained in DNA sequences selected from
CC   ACD18988-ACD19242 (being 255 E.coli 0157 DNA sequences which are not
CC   found in E.coli K12), a recombinant DNA construction comprising the DNA
CC   above and a method for detecting E. coli 0157:H7 (ATCC 43895) in a sample
CC   (or distinguishing between 0157 and K12) using a probe derived from one
CC   of the 255 sequences. The DNA sequences are useful in detecting E. coli
CC   O157:H7 in a sample, for the early diagnosis of humans and livestock
CC   infected with O157:H7, and in designing diagnostic probes which can be
CC   used to distinguish strain O157:H7 from strain K12 using molecular
CC   techniques. The present sequence is one of the 255 E. coli 0157:H7 DNA
CC   sequence (termed OZID_1-OZID255). Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from the USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030023075 (Updated on 27-OCT-2003
CC   to standardise OS field)
XX
SQ   Sequence 14187 BP; 4451 A; 2859 C; 2282 G; 4595 T; 0 U; 0 Other;

  Query Match             99.8%;  Score 781.4;  DB 9;  Length 14187;
  Best Local Similarity   99.9%;  Pred. No. 1.9e-153;
  Matches  782;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13922 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 13863

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13862 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 13803

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13802 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 13743

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13742 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 13683
```

```
Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13682 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 13623

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13622 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 13563

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      13562 TATTATGTTATAGATAACAAAAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 13503

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13502 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 13443

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13442 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 13383

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13382 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 13323

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13322 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 13263

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13262 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 13203

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13202 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 13143

Qy        781 TGA 783
              |||
Db      13142 TGA 13140
```

```
RESULT 5
AEC31468
ID   AEC31468 standard; DNA; 18772 BP.
XX
AC   AEC31468;
XX
DT   20-OCT-2005  (first entry)
XX
DE   E. coli type O24 O-antigen gene cluster.
XX
KW   O-antigen; gene expression; DNA detection; gene; ds.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1096. .1875
FT                   /*tag=  a
FT                   /note= "Orf1. Specifically claimed in Claim 3"
FT   CDS             3584. .4204
FT                   /*tag=  b
FT                   /note= "Orf2"
FT   CDS             4209. .5246
FT                   /*tag=  c
FT                   /note= "Orf3"
FT   CDS             5249. .6508
FT                   /*tag=  d
FT                   /note= "Orf4"
FT   CDS             6505. .7692
FT                   /*tag=  e
FT                   /note= "Orf5"
FT   CDS             9380. .10387
FT                   /*tag=  f
FT                   /note= "Orf6"
FT   CDS             10411. .11463
FT                   /*tag=  g
FT                   /note= "Orf7"
FT   CDS             11479. .12672
FT                   /*tag=  h
FT                   /note= "Orf8"
FT   CDS             16255. .17325
FT                   /*tag=  i
FT                   /note= "Orf9. Specifically claimed in Claim 3"
XX
PN   CN1563037-A.
XX
PD   12-JAN-2005.
XX
PF   19-APR-2004; 2004CN-00019022.
XX
PR   19-APR-2004; 2004CN-00019022.
XX
PA   (TIAN-) TIANJIN BIOCHIP TECH CO LTD.
XX
PI   Wang L,  Wang W,  Feng L;
XX
DR   WPI; 2005-316281/33.
XX
PT   New nucleotide specific to the O-antigen of O24 type Escherichia coli,
PT   for detecting and identifying Escherichia coli type O24 in a human body.
XX
PS   Claim 1; SEQ ID NO 1; 35pp; Chinese.
XX
CC   The invention relates to a nucleotide sequence (SEQ ID No:1) which is
CC   specific to the O-antigen of Escherichia coli O24. The nucleotide
CC   sequence represents the complete sequence (18772bp) of a gene cluster for
CC   controlling the synthesis of O-antigen in Escherichia coli O24. The
CC   nucleotide sequence can have one or several inserted, deleted or
CC   substituted bases, whilst still retaining its function. Also described is
CC   a PCR primer for the glycosyltransferase (galF) gene and a PCR primer for
CC   the oligosaccharide unit treatment (gnd) gene, both of which are
CC   contained within the E. coli type O24 O-antigen gene cluster. The
CC   invention also discloses a method for detecting and identifying E. coli
CC   type O24 in humans. This sequence represents the E. coli type O24 O-
CC   antigen gene cluster of the invention.
XX
SQ   Sequence 18772 BP; 5911 A; 2966 C; 3871 G; 6024 T; 0 U; 0 Other;

  Query Match            69.5%;  Score 544;  DB 14;  Length 18772;
  Best Local Similarity  82.2%;  Pred. No. 7e-104;
  Matches  625;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||| |||||||| ||||||| | |||||||| || || ||||||||   |||||||||
Db       1096 ATGAATAAAGAAACTGTTTCAGTTATTATGCCTGTCTATAATGGGGCTCGAACTATAATT 1155

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||| |||||| ||||||| |||||||| |||||||| ||||||||||||
Db       1156 TCATCAGTAGAATCGATTATAAATCAATCATATCAAGACTTTGTTTTATATATCATTGAC 1215

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              |||||||||||||||||||||||||||||    | |||||| |||||||||| |||
Db       1216 GATTGTAGCACCGATGATACATTTTCATTAATTTTTAATCGATATAAAAACAATCAAAAA 1275

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||| || ||||| ||  |  | |||||||||| || |||||| |||||  |||||||
Db       1276 ATAAGGATGTTGCGGAATGACTCTAATTTAGGTGTGGCGGAAAGTAGAAATGTTGGAATA 1335

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||   || || || |||||||||||||||||||||||||||||||| |||||| ||||
Db       1336 AAAGCAGCAACAGGAAAATATATTTCTTTTTGTGATGCGGATGATTTATGGCACCAGAAT 1395

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              |||||||| || ||||||||| | || || |||||||||| ||||||||||||||||||
Db       1396 AAATTAGAACGACAAATCGAATTATTCAACAATGAATGTGTTGATGTGGTATGTTCTAAT 1455
```

```
Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              ||||||||  |  |||  |||  |  ||  ||  ||||  ||  |||||  ||  |||||||  |  |||
Db       1456 TATTATGTGGTGAATATTAATGGGGATGTTATTGGTGAGGTTAAAGCACCTCATATCATA 1515

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              | |||  ||||||||||| ||||||||   | ||||||||||| ||  ||  ||  ||||||||
Db       1516 ACTTACAGAAAAATGCTTATGAAAAATCATATAGGGAATTTAACTGGTATTTATAATGCT 1575

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||| |||||||||||||| |||||  |||||||||||||| ||||||  ||
Db       1576 AACAAATTGGGTAAATTTTATCAAAAAAATATTGGCCACGAGGATTATTTAATGTGGATG 1635

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||  |||||  ||  |||||||| | |||  |||||||||||||| |||||  ||  |||||||||
Db       1636 GAGATAATAAACAAAACAAACGTTGCCATTTGTATTCAAGACAATCTCGCTTATTACATG 1695

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||  |||  |||||||||  ||  | |||||||  |  ||  ||||||||||||  ||||  ||||
Db       1696 CGGACAAGTAATTCACTCTCTGCTAATAAAGTGAAGACTGCAAAATGGATGTGGAATATA 1755

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||  ||   ||||||||  ||  |||  |  |||  |  |||||||  ||||||   |  |||||||
Db       1756 TATAGACAATGCTTACATTTATCTTTTACTAAATCTTTATATTGTTTTTTTATGTATGCT 1815

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAA 760
              |  |||||  ||  ||  ||  |  ||||  ||||  ||  |||  |
Db       1816 TTTAATGGGGTTATTAAGGAGATAATGCATTTACCATTGA 1855
```

RESULT 6
ACA29283
ID ACA29283 standard; DNA; 747 BP.
XX
AC ACA29283;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #10940.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Clostridium difficile.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU25413.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 17153; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 747 BP; 314 A; 74 C; 125 G; 234 T; 0 U; 0 Other;

Query Match   23.2%; Score 181.8; DB 8; Length 747;
Best Local Similarity 54.6%; Pred. No. 1.8e-28;
Matches 407; Conservative   0; Mismatches 332; Indels   6; Gaps   2;

```
Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              |||||  ||  |    ||  ||  |||||||  |||  |  ||  |||   |  |||  |||
Db          1 ATGAAGAAGAATTTAGTATCTATAATTACTCCCATGTATAATTCTGAAAAATTTATTGAA 60

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
                ||  |  ||  |||||  |  ||  |  |||  ||||||||||  |  |    |||  ||  |||||
Db         61 GCAACCATAAAATCAGTATTAAACCAAACTTATCAAGAATGGGAAATGTTAATTATTGAT 120

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              |||||    ||  |||  |||    |    ||  ||||    |          ||  |  ||    |  |
Db        121 GATTGCTCAACAGATAATAGTCCTAATATAGTCAAATCTTATATGCAACAGGATAGTAGA 180

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||  |    |    |    ||||  |||    |||||  |  |    |  ||||||   |  | ||
Db        181 ATAAAATGTATAAAGACTGAGACTAATAAGGGTGTCTCTAATGCTAGAAATTTAGCACTA 240

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              | |||  ||  ||  |||  |||  ||||||  |||    |||||          |||  |    |
Db        241 AGTAAGGCAACAGGACAATTTATAGCTTTTTTAGATAGTGATGACCAATGGAATAGTAGT 300

Qy        301 AAATTAGAGCGTCAAATCGA---AGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCT 357
              ||  |||||     |||  |  |     |||||  |  |||||  |  ||||   |     |   |||
Db        301 AAGTTAGAAAAACAAGTAAATTTTATGTTAGAAAATGACTATGTAATTTCATTTACTTCA 360

Qy        358 AATTATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTG 417
              ||  |    |  ||  ||  |||  |||  |          |  |  |  ||  |  |||  |  |
Db        361 TATGAACTGATGGATGAAAATGATAAAAAATTAAACAAAGTAATAAAAGTTCCACCTAAT 420

Qy        418 ATAAATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAAT 477
              ||  |||||||||  |  |  |  |     |||    |  ||||    |  |  ||  |    |  | ||
Db        421 GTAGATTATAGAAGATTATTAAAAGGAAATATCTTAGGATGTCTTACTGTTGTGATTGAT 480

Qy        478 GCCAACAAATTGGGTAAGTTTTATCAAAAAAAGATTG---GTCACGAGGATTATTTGATG 534
                  |||||||  |              ||  |    ||    |  ||  ||  ||||||  |  |
Db        481 AAGTCTAAATTGGATTTTGAAATAAGAATGAGTGGTGTAAGACATGAAGATTATGTTTTA 540

Qy        535 TGGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTAT 594
```

```
Db        541 TGGTTATCGATATTGAAAAAAGGGCATATTGCTCATGGAATAAATGAAGTTCTAGCATTA 600

Qy        595 TACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGG 654

Db        601 TATAGAAAATCTAGCAATTCCCTAAGTGGTAATAAAATAAAAGCTGCTATGTGGACTTGG 660

Qy        655 AGTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTA 714

Db        661 AATATATATAGAAATATAGAAAAAATACCTTTATACAAAGCTATATACTACTTTATTAAT 720

Qy        715 TATGCTTCAAATGGAGTCATGAAAA 739

Db        721 TATGGAATAAACGGAATCAAAAAAA 745


RESULT 7
ACA28974
ID   ACA28974 standard; DNA; 774 BP.
XX
AC   ACA28974;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Prokaryotic essential gene #10631.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Clostridium difficile.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU25104.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 16844; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC   prokaryotic essential genes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 774 BP; 306 A; 74 C; 152 G; 242 T; 0 U; 0 Other;

  Query Match             20.8%;  Score 162.8;  DB 8;  Length 774;
  Best Local Similarity   52.9%;  Pred. No. 1.6e-24;
  Matches  397;  Conservative    0;  Mismatches  347;  Indels    6;  Gaps    2;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60

Db         25 ATGAATGAGCCTTTAGTATCGATAATTACTCCTGTTTATAATTCAGAAGAGTTTTTATCA 84

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120

Db         85 GAGACGATAAAGTCAATCCAAAATCAAACTTATAAAAATTGGCAGCTTTTATTAGTGGAT 144

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180

Db        145 GATTGTTCTAAGGACAATTCAAGTGAAATTATAAAAAGTTTTCGAAAAGAAGATGCTAGA 204

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240

Db        205 ATAAAATATATAAAGCTTGAAAAAAATTCTGGAGCTGCAGTATCTAGAAATGTTGGTATT 264

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300

Db        265 AAGAATGCGGAAGGTCGATTTATTGCATTTGTAGATAGTGATGATTTATGGGATAGTAGA 324

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGT---TCT 357

Db        325 AAATTAGAGATTCAGATAGAATATATGTTAAAAGAAAATGTAGGTTTTTCTTTTACATCT 384

Qy        358 AATTATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTG 417

Db        385 TATAGATATATGAGACAAGATGGGTCAAAAACGAATAAGGTTGCAAGAGCTCCTAAGAAA 444

Qy        418 ATAAATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAAT 477

Db        445 ATTGACTATGAAGGATTACTTAGAAATACTATAATAGGATGTTCTACAGTGGTCATTGAT 504

Qy        478 GCCAACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACG---AGGATTATTTGATG 534

Db        505 AAGGAGATTGTAGGAGAGTTCAGCATGCCTTTGGTAAGAAGAGGACAAGATACTGCAACC 564

Qy        535 TGGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTAT 594

Db        565 TGGTTACAGTTACTTAAAAAGGAAAAATATGCATATGGGATTCAAGAAGATTTAGTGAAT 624

Qy        595 TACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGG 654

Db        625 TACAGATTAGTTGGAAATTCTATATCAAGTAACAAGATAAAAGCATTAAAACGAACTTGG 684

Qy        655 AGTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTA 714

Db        685 AATACTTATAGAAATGTTGAAAATTTAAGTTTACCAAAAAGTCTATATGTATTTTGTTTC 744

Qy        715 TATGCTTCAAATGGAGTCATGAAAAAAATA 744

Db        745 TATGTTTTTAATGCCATTAAAAAAAGAGTA 774
```

```
RESULT 8
AEA88828
ID   AEA88828 standard; DNA; 12220 BP.
XX
AC   AEA88828;
XX
DT   11-AUG-2005  (first entry)
XX
DE   Escherichia coli O18 O-antigen biosynthetic gene cluster.
XX
KW   Microorganism detection; O-antigen biosynthetic gene cluster;
KW   lipopolysaccharide antigen; gene; ds.
XX
OS   Escherichia coli; O18.
XX
FH   Key             Location/Qualifiers
FT   CDS             1415. .2434
FT                   /*tag= a
FT                   /product= "Orf1 protein"
FT   CDS             2800. .3171
FT                   /*tag= b
FT                   /product= "Orf2 protein"
FT   CDS             3407. .4567
FT                   /*tag= c
FT                   /product= "Wsx protein (orf3)"
FT   CDS             4551. .5723
FT                   /*tag= d
FT                   /product= "Orf4 protein"
FT   CDS             5720. .6526
FT                   /*tag= e
FT                   /product= "Orf5 protein"
FT   CDS             6516. .7745
FT                   /*tag= f
FT                   /product= "Wsy protein (orf6)"
FT   CDS             8021. .8755
FT                   /*tag= g
FT                   /partial
FT                   /product= "Orf7 protein"
FT                   /note= "No start codon given"
FT   CDS             8800. .9846
FT                   /*tag= h
FT                   /product= "Orf8 protein"
FT   CDS             9843. .10631
FT                   /*tag= i
FT                   /product= "Orf9 protein"
XX
PN   CN1554759-A.
XX
PD   15-DEC-2004.
XX
PF   22-DEC-2003; 2003CN-01117863.
XX
PR   22-DEC-2003; 2003CN-01117863.
XX
PA   (UYNA-) UNIV NANKAI.
XX
PI   Wang L,  Kong Q,  Feng L;
XX
DR   WPI; 2005-215827/23.
XX
PT   O-antigen specific nucleotide of E. coli 081 type.
XX
PS   Claim 1; SEQ ID NO 1; 29pp; Chinese.
XX
CC   The invention relates to the O-antigen biosynthetic gene cluster of
CC   Escherichia coli O81 (AEA88828) and to variants thereof. The O-antigen of
CC   E. coli O81 is a lipopolysaccharide antigen. The invention also relates
CC   to PCR primers (AEA88829-AEA88830) for the amplification of the gnd gene
CC   of the O-antigen gene cluster, oligonucleotides derived from the
CC   glycosyltransferase and oligose unit processing genes of the O-antigen
CC   gene cluster, and a method for detecting E. coli O81 in the environment
CC   and in the human body using oligonucleotides of the invention. The
CC   present sequence represents the specifically claimed sequence of the E.
CC   coli O81 O-antigen biosynthetic gene cluster.
XX
SQ   Sequence 12220 BP; 3952 A; 1820 C; 2307 G; 4141 T; 0 U; 0 Other;

  Query Match             20.8%;  Score 162.8;  DB 14;  Length 12220;
  Best Local Similarity   52.3%;  Pred. No. 2.4e-24;
  Matches  387;  Conservative    0;  Mismatches  347;  Indels    6;  Gaps    1;

Qy          8 AAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAG 67
              | |||||  || ||  ||||| |||||    |  ||||  ||      |||    |
Db       9853 ATGAAACAGTATCTATAATCATGCCTACGTTTAATGCTGCATCTTATATTTCGTATAGCA 9912

Qy         68 TAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTA 127
              |   ||  | ||  | ||||| ||| | ||||    ||| |||||  || || ||| ||
Db       9913 TCATTTCTGTAATTGAACAATCATATAATGATTGGCATTTATATATAATAGATGATGGTT 9972

Qy        128 GCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAA 187
               || ||  ||||          |||     | |           | | ||| |  ||||| |
Db       9973 CAACTGACAATACCGCACATGAAATATCTAATTTT------ACCGATCTGCGAATAACAT 10026

Qy        188 TATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGG 247
               |           | |||   || |||||| |   | |||||   || ||  ||||||
Db      10027 ATGTAATGCTTCCTATCAATAGTGGAGTTGCAAATGCTAGAAATATAGGTATTAAAATGG 10086

Qy        248 CCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAG 307
              | |  ||||||| |||| | |||  ||||   |||||  | ||| |  |  |||||||||
Db      10087 CAAAAGGGAAATTTATTGCGTTTCTTGATAGTGATGACATTTGGGAAAAAGAAAAATTAT 10146

Qy        308 AGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATG 367
                 ||||||  |  |  | || | || ||  |  ||    |   | ||     |
Db      10147 CCTTGCAAATCCCTTTACTCGAAAAAGGATATGATGTAGTTTGTTCAAATTATGCCATTT 10206

Qy        368 TTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATA 427
              ||| | || |    | ||| ||  |  ||         | ||| || ||   ||| | |||
Db      10207 TTAAAAATGAACCAAAAAAAAATAATCGGTTCTCGAACTGCACCACAACAGATCACCTATT 10266

Qy        428 GAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAAT 487
                || |       || |  || || ||||| || || || |||||      |  |
Db      10267 CGCGTTTGTTACGCGGTAATTGTATCGGCAATTTAACGGGGATTTATAACCGAGAACGAC 10326

Qy        488 TGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGGAAATAA 547
              | ||||| |  |  ||  || |  ||| || || || ||| | |||||| |  |  |
Db      10327 TAGGTAAATGCTTACAGGAAGAATGTGGACATGAAGACTATCTAATGTGGTTACAGTTGG 10386

Qy        548 TTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCAA 607
              | ||      |  | | |||  ||| || ||  | ||| | ||    ||     | |
Db      10387 TAAAGCTTTCTGGTAGCGCTTATTGCATACAGAAAAATTTAGCCCGGTATCGAGTTACGG 10446

Qy        608 ATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATATATAGAG 667
              |  |||||  ||||    |||||| | ||||| |||   |||  |||| |||||||  |
Db      10447 AACATTCAGTATCATCAAATAAATTAAAAGCAGCATTTTGGCAATGGAATATATATCGTA 10506

Qy        668 AACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCTTCAAATG 727
              || |  | |  |    ||   |  |  |||||||  |  |     ||| |||   | |
Db      10507 AAAAACTCAAGCTTAATATTTTGAGCAGTATATATTACTGGTGTAATTATTCTTTTTACG 10566

Qy        728 GAGTCATGAAAAAAATAACA 747
              | | ||| | |   ||| |
Db      10567 CATTTATGCACATGTTAAAA 10586


RESULT 9
AAX13062/c
ID   AAX13062 standard; DNA; 4915 BP.
XX
AC   AAX13062;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:125.
XX
```

KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.
XX
PN WO9850555-A2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US008985.
XX
PR 06-MAY-1997; 97US-0044031P.
PR 16-MAY-1997; 97US-0046655P.
PR 14-NOV-1997; 97US-0066009P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kunsch CA, Dillon PJ, Barash SC;
XX
DR WPI; 1999-045171/04.
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides -
PT used to develop products for the detection of Enterococcus and for use in
PT vaccines for prevention or attenuation of Enterococcus infection.
XX
PS Claim 1; Page 759-761; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence of
CC Enterococcus faecalis in samples. They can also be used for diagnosing
CC Enterococcal infection in an animal and monitoring progression of
CC disease, and for identifying agents which can be used to modulate the
CC growth or pathogenicity of Enterococcus faecalis, or another related
CC organism, in vivo or in vitro. In particular the polypeptides encoded by
CC the Enterococcus faecalis nucleotide sequences can be used in vaccines to
CC prevent or attenuate an Enterococcal infection
XX
SQ Sequence 4915 BP; 1465 A; 950 C; 750 G; 1740 T; 0 U; 10 Other;

Query Match 17.9%; Score 140.4; DB 2; Length 4915;
Best Local Similarity 50.8%; Pred. No. 9.9e-20;
Matches 361; Conservative 0; Mismatches 346; Indels 3; Gaps 1;

```
Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||| | | || |   |||||||| ||||||||| | |||||||  | ||||  | ||
Db       4211 ATGGAAACAGCACTTGTTTCAATTATTATGCCCATGTACAATGCTGGCAAATTTTTATCA 4152

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
                |||  | |||    || |  | ||| | |||||  |||  |  || |   |||||||
Db       4151 AAATCTATTGAAAGCGTTCTGGAACAAACCTATCAGAATTGGGAATTATTACTCATTGAT 4092

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||| | || |  |||||| |  ||     |                ||||  ||   | |
Db       4091 GATGGAAGTAAAGATGATTCTATTGATATTGCRCTTGCCTTTATGGAAAAAGATTCAAGA 4032

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||     || ||   ||  |    ||  | ||  ||||| | |   ||||| | |||||
Db       4031 ATCTTTTTACTGAAAAATGAACAGAACATGGGCATTGCAAAGACCAGAAATAAAGGAATT 3972

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              |||    | | |||  |||||||| | ||||  |||   |||||||| ||||     ||
Db       3971 GAAGCTTCAAAGGGACAATATATTGCCTTTTTAGATAGTGATGATTTATGGCTTCCTAAT 3912

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||   |||| || |   | |  |  |  ||    |      |   ||||  |  |
Db       3911 AAATTAGAAGTTCAAGTCAAGTGGATGGAAGAAAAAAAACTTTTATTTACATGTAGTTCT 3852

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATG---TG 417
              |||| |||     || | |'| | || |||  |   ||   ||||  |  | | |
```

```
Db       3851 TATTTTGTATGTAATGAAAATGGCAACATTACTCATGAGCGTAATTTTAGTGAGGGACCA 3792

Qy        418 ATAAATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAAT 477
                ||  |||  | |  |  | |  | |||    || ||    || ||||   |      |
Db       3791 CAAACCTATCAAGATCTTTTAAAAACAAATACTATCGGCTGCTTAACAGTCGTAGTCGAA 3732

Qy        478 GCCAACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGG 537
                | ||     |          ||          |  |    || || || |||    |  |||
Db       3731 TCTAATCTGCTTAAACGCCATTTAATGCCGGATTTAAAACATGAAGACTATGCTACTTGG 3672

Qy        538 CTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTAC 597
                |  | ||  | ||  ||| ||||   |    || |||| | || ||  | ||   |||
Db       3671 TTAAACATTTTGAAAGAAATAAATACGGTCTATTTTATTAACGAAAAATTAGCCATTTAT 3612

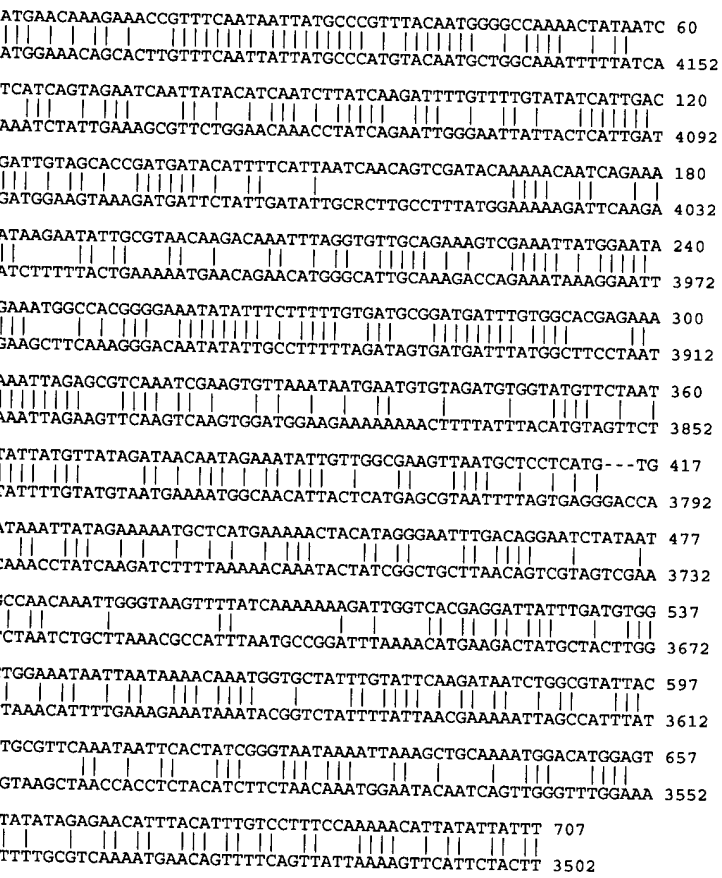
Qy        598 ATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGT 657
                  ||  |  ||   |||   ||| |||   ||  |    |  |||   ||||
Db       3611 CGTAAGCTAACCACCTCTACATCTTCTAACAAATGGAATACAATCAGTTGGGTTTGGAAA 3552

Qy        658 ATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTT 707
                || |   |  || ||  ||| || ||  ||   ||||  | ||  || ||
Db       3551 ATTTTGCGTCAAAATGAACAGTTTTCAGTTATTAAAAGTTCATTCTACTT 3502
```

RESULT 10
ABS98857/c
ID ABS98857 standard; DNA; 4915 BP.
XX
AC ABS98857;
XX
DT 18-DEC-2002 (first entry)
XX
DE Enterococcus faecalis contig sequence #125.
XX
KW Computer readable medium; Enterococcus faecalis; microbe; growth;
KW pathogenicity; vaccine; resistance; Enterococcal infection; commercial;
KW therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;
KW biochip technology; antibacterial; modulator of nucleic acid expression;
KW contig; ds.
XX
OS Enterococcus faecalis.
XX
PN US2002120116-A1.
XX
PD 29-AUG-2002.
XX
PF 04-MAY-1998; 98US-00070927.
XX
PR 04-MAY-1998; 98US-00070927.
XX
PA (KUNS/) KUNSCH C A.
PA (DILL/) DILLON P J.
PA (BARA/) BARASH S.
XX
PI Kunsch CA, Dillon PJ, Barash S;
XX
DR WPI; 2002-750065/81.
XX
PT Computer readable medium having recorded on it a Enterococcus faecalis
PT nucleotide sequence useful for detecting diseases related to Enterococcus
PT infections in animals.
XX
PS Claim 1; Page; 119pp; English.
XX
CC The present invention relates to a new computer readable medium with an
CC Enterococcus faecalis nucleotide sequence. The invention is useful to
CC diagnose the presence of E.faecalis in a sample or determining the
CC presence of a specific microbe in a sample. The invention is also useful
CC for modulating the growth or pathogenicity of E.faecalis, in a vaccine to
CC confer resistance to Enterococcal infection, for commercial, therapeutic
CC and industrial purposes, and for fermenting a particular sugar source or
CC to produce a particular metabolite. The invention is useful for detecting
CC diseases related to Enterococcus infections in animals, and for detecting
CC E.faecalis using biochip technology. The present nucleic acid sequence

CC represents an Enterococcus faecalis contig DNA sequence of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at http.sequdata.uspto.gov
XX
SQ Sequence 4915 BP; 1465 A; 950 C; 750 G; 1740 T; 0 U; 10 Other;

Query Match 17.9%; Score 140.4; DB 6; Length 4915;
Best Local Similarity 50.8%; Pred. No. 9.9e-20;
Matches 361; Conservative 0; Mismatches 346; Indels 3; Gaps 1;



RESULT 11
ADV98191
ID ADV98191 standard; DNA; 30248 BP.
XX
AC ADV98191;
XX
DT 24-FEB-2005 (first entry)
XX
DE Shigella boydii DNA encoding a glycosyltransferase Seq 1.
XX
KW O-antigen; glycosyltransferase; microorganism detection; ds.
XX
OS Shigella boydii.
XX
FH Key Location/Qualifiers
FT misc_feature 3311. .4228
FT /*tag= a
FT /label= ORF3
FT misc_feature 4664. .5869
FT /*tag= b
FT /label= wzx
FT misc_feature 5851. .6744
FT /*tag= c
FT /label= ORF6
FT misc_feature 6741. .7856
FT /*tag= d
FT /label= ORF7
FT misc_feature 7853. .8935
FT /*tag= e
FT /label= wzy
FT misc_feature 12131. .12934
FT /*tag= f
FT /label= ORF10
XX
PN CN1439645-A.
XX
PD 03-SEP-2003.
XX
PF 20-JAN-2003; 2003CN-00100356.
XX
PR 20-JAN-2003; 2003CN-00100356.
XX
PA (UYNA-) UNIV NANKAI.
XX
PI Wang L, Guo H, Feng L;
XX
DR WPI; 2004-157378/16.
XX
PT Nucleotide with O-antigen idiocrasy against Ball's and shiga's-8 and
PT colibacillus 0143.
XX
PS Claim 1; SEQ ID NO 1; 35pp; Chinese.
XX
CC This invention relates to a novel nucleic acid molecule specific to the O
CC -antigen of Shigella boydii 8 and colibacillus 0143. Specifically, this
CC DNA sequence comprises 15124 bases. The present invention describes this
CC sequence as one that contains a glycosyltransferase gene and further
CC provides oligonucleotides such that it can be amplified. Accordingly, the
CC primers provided herein can be used to detect Shigella boydii 8 and
CC colibacillus 0143 in the human body and also where it is present in the
CC environment. This polynucleotide sequence is the Shigella boydii 8 DNA
CC that is specific to the 0-antigen, given in an exemplification of the
CC invention.
XX
SQ Sequence 30248 BP; 9696 A; 4220 C; 5968 G; 10364 T; 0 U; 0 Other;

Query Match 17.7%; Score 138.4; DB 13; Length 30248;
Best Local Similarity 49.9%; Pred. No. 3.3e-19;
Matches 377; Conservative 0; Mismatches 376; Indels 3; Gaps 1;



```
Qy        245 TGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAAT 304
Db      12381 AAGCTAAAGGTCGGTTTATCGCTTTTTTAGATAGTGATGATTTTTGGCATAAAGAGAAAA 12440

Qy        305 TAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATT 364
Db      12441 TAAAAAAACAGATAGACTTTATGCTTTCAAATAATTATGCTTTCACTTATACCTCATATC 12500

Qy        365 ATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATT 424
Db      12501 AAAAGGTTAGTTTTAGCGGGAAACTTTTATCCCAAATTAATCCTGTTTCTAAAGTTAATT 12560

Qy        425 ATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACA 484
Db      12561 ATTCTGAATTATTAAAGACTAATGTAATAGGTTGCTTGACAGCTGTATATGATACTTCAT 12620

Qy        485 AATTGGGTAA---GTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGG 541
Db      12621 CTCTTGGTAAAATGTTTATGCCAACAATCAGGAAAAGACAGGACATGGCACTATGGCTGA 12680

Qy        542 AAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGC 601
Db      12681 GCATTTTAGATAAAATTGATTATGCATATTGTTTGAACGAAATACTGGCCTACTATCGAG 12740

Qy        602 GTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATAT 661
Db      12741 TAGGCACCGATACACTCTCGTCAAATAAATTAAAAATAATATTTTCACAATGGGCCTTTT 12800

Qy        662 ATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCTT 721
Db      12801 ACAGAAAATATTTAAAGTTTGGTCTGTTTAAATCTCTTTATTATTTCTCCCACTATGCTG 12860

Qy        722 CAAATGGAGTCATGAAAAAAATAACACATTCACTAT 757
Db      12861 TTAATGCTGTTGCCAAACATCATATAAATCCGCTGT 12896
```

RESULT 12
ADN10409
ID ADN10409 standard; DNA; 10812 BP.
XX
AC ADN10409;
XX
DT 17-JUN-2004 (first entry)
XX
DE O-antigen sequence, SEQ ID 1.
XX
KW O-antigen; detection; ds.
XX
OS Shigella boydii.
XX
PN CN1438230-A.
XX
PD 27-AUG-2003.
XX
PF 20-JAN-2003; 2003CN-00100538.
XX
PR 20-JAN-2003; 2003CN-00100538.
XX
PA (UYNA-) UNIV NANKAI.
XX
PI Wang L, Guo H, Feng L;
XX
DR WPI; 2003-845896/79.
XX
PT Nucleotide.
XX
PS Claim 1; SEQ ID NO 1; 31pp; Chinese.
XX
CC The present invention relates to the nucleotide sequence for the O-
CC antigen of Shigella boydii (ADN10409). The invention also discloses a
CC method for testing the sequence and detecting Shigella boydii 15,
CC Shigella dysenteriase 2 and colibacillus 0112 in the human body and the
CC environment.
XX
SQ Sequence 10812 BP; 3567 A; 1592 C; 2045 G; 3608 T; 0 U; 0 Other;

  Query Match   16.2%; Score 126.6; DB 11; Length 10812;
  Best Local Similarity 50.7%; Pred. No. 8.3e-17;
  Matches 384; Conservative   0; Mismatches 364; Indels   9; Gaps   3;

```
Qy         10 GAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTA 69
Db       7656 GAATTAGTTTCAATTATAATGCCTTGTTTTAATAATGAAGCAACTATAAAAGAAAGTATT 7715

Qy         70 GAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGC 129
Db       7716 AATAGCGTACTATCTCAAAGTTATAAATACTGGGAGTTAATTATTATCAATGACAACTCA 7775

Qy        130 ACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATA 189
Db       7776 ACAGATGATTCACAAAATATTATTAACGACTATGTGAATAGTGATGAAAGAATAAAACTG 7835

Qy        190 TTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCC 249
Db       7836 ATTACTAATGAAAAAAATATCGGAGCTGGAGGTAGTAGAAATAAAGGAATTGAAATTGCA 7895

Qy        250 ACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAG 309
Db       7896 AATGGAAGATTCATTGCTTTTTTAGATGCAGATGATGTTTGGTTGCCAGAAAAACTTGCA 7955

Qy        310 CGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTT 369
Db       7956 AAACAAATAAAGTTTATGAGAGATGGTAATTATGCACTGACATATAGTTATTATCAAAAG 8015

Qy        370 ATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGA 429
Db       8016 TTTGATTCAACTGGA---TTGGGAGCGATCGTCCGGTCACCTGCGACAACAACTTATAGG 8072

Qy        430 AAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTG 489
Db       8073 AAATTGTTATACAGTAATGTTATTGGATGTCTAACAGGAATGTATGATACTGATTTTTTG 8132

Qy        490 GGTAAGTTTTATCAAAAAAAGATTGGT---CACGAGGATTATTTGATGTGGCTGGAAATA 546
Db       8133 GGTAAATGCTATATGCCTTTAATTCGTAAACGACAAGATATGGGGCTTTGGTTGTCAATA 8192

Qy        547 ATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCA 606
Db       8193 TTGAAAAAGTGTAATAAAGCATATTGCTTACCTGAAGTTTTGGCATATTATAGAATTG-- 8250

Qy        607 AATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATATATAGA 666
Db       8251 -ACTCAGGAATGACAAAAAATAAACTTGATGCGGCAAAATATCAGTGGCGATTTTATAGG 8309

Qy        667 GAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCTTCAAAT 726
Db       8310 GATGTAACTAAACTTAATTTGATTCAATCAAGTTGGTATTTTGTTTGGTATTCAGTTTTA 8369

Qy        727 GGAGTCATGAAAAAAATAACACATTCACTATTAAGGA 763
Db       8370 GGTTTGGTAAAGTACAAAAAATAATAAGGAGGAGTGA 8406
```

RESULT 13
ACA34280
ID ACA34280 standard; DNA; 753 BP.
XX
AC ACA34280;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #15937.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.

XX
OS Haemophilus influenzae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU30410.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 22150; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 753 BP; 292 A; 96 C; 111 G; 254 T; 0 U; 0 Other;

Query Match 14.9%; Score 116.8; DB 8; Length 753;
Best Local Similarity 49.1%; Pred. No. 6.5e-15;
Matches 370; Conservative 0; Mismatches 377; Indels 7; Gaps 2;

```
Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||| |    |  | ||  ||||| ||||| ||||| |||| |           | |
Db          1 ATGAATATGCCACTCATTAGTATAATAATGCCTGTTTATAATGCAGAATGTTACCTTAAT 60

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
                 |   |  |||     || ||||||||||||||  |  |||  ||   | | |||||
Db         61 CAAGGTATTTTATCTTGCTTAAATCAATCTTATCAAAACATTGAATTAATTTTAATTGAT 120

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||| ||   || ||| | |   ||  | |||| || |           | ||  || ||
Db        121 GATGGTTCAACAGATAAAAGTATTGAAATAATAAATAACATCATTGATAAAGATAAGCGG 180

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              | | | | ||   || |   || ||| |||  |||||      |||||   ||  ||
Db        181 GTTAAACTTTTCTTTACCCCAACTAATCAAGGGCCTGCAGCGGCAAGAAATATAGGGTTA 240

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||  ||    ||  | |||||| | || |  |||  ||||||||          | |
Db        241 GAAAAAGCACAAGGTGACTATATTACATTCTTAGATAGTGATGATTTTATTGCAAATGAT 300

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||  ||| | |  | |  |          | |  | | | ||  |  | |  ||
Db        301 AAATTAGAAAAACAACTTAATTTTATGTTGCAAAATCATCTTGTAATGACGCATGGCAAC 360

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||  | |  |||    | |||||    |     |||||   ||  | |   ||
Db        361 TATGCTTTCTGTGATTTGGAAGGAAATCAGATAAAATTAGTTACGACTAGTAAAAAAATT 420

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              |||||  ||   |  |          ||| | |  |  |  |||| | |  | |
Db        421 GATTATTTAACCTTATTACAAGGTAACCAATTTAAAATAATGACTGTACTTGTTGAAAGA 480

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTA-----TTTGATGT 535
              |    |    ||    || |||  ||||| |||||     | | |
Db        481 GAATCTATTAAATTACTTAGGTTTCCTAATATTAAACACGAAGATTACGCCTTCTTTTTG 540

Qy        536 GGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATT 595
              |  ||   | ||  | | |||||  | |||| |    ||  ||  |   | |
Db        541 GATTGTTTAAAAGAAGTTAAACAATCTATACTATATAGCCATCAGGCAAGTTCATTTGTA 600

Qy        596 ACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGA 655
              | |  || | ||| |  |    | ||   |||||| |||||  |||||  ||||||| |
Db        601 AGAATCGGTAAAGTCAGCG--TTTCATCAAATAAATTTAAAAGTGCAATCTGGACATTTA 658

Qy        656 GTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTAT 715
              ||| |||   ||     |  |     |  |||| || | |||||||||  |  | |
Db        659 ATATCTATTTTAAAAGAGAGAAACTAGGTGTAGTAAAATCAATTTATTATTTTTATCCTCT 718

Qy        716 ATGCTTCAAATGGAGTCATGAAAAAAATAACACA 749
              |||| |  |||||| | || ||| | | | | |
Db        719 ATGCCTATAATGGATTTATTAAATACAAGAAATA 752
```

RESULT 14
AAT42063_09
Continuation (10 of 19) of AAT42063 from base 900001 (Haemophilus influenzae complete ger
WP Sequence split into 19 fragments LOCUS AAT42063 Accession Aat42063

	Fragment Name	Begin	End
WP	AAT42063_00	1	110000
WP	AAT42063_01	100001	210000
WP	AAT42063_02	200001	310000
WP	AAT42063_03	300001	410000
WP	AAT42063_04	400001	510000
WP	AAT42063_05	500001	610000
WP	AAT42063_06	600001	710000
WP	AAT42063_07	700001	810000
WP	AAT42063_08	800001	910000
WP	AAT42063_09	900001	1010000
WP	AAT42063_10	1000001	1110000
WP	AAT42063_11	1100001	1210000
WP	AAT42063_12	1200001	1310000
WP	AAT42063_13	1300001	1410000
WP	AAT42063_14	1400001	1510000
WP	AAT42063_15	1500001	1610000
WP	AAT42063_16	1600001	1710000
WP	AAT42063_17	1700001	1810000
WP	AAT42063_18	1800001	1830121

Query Match 14.9%; Score 116.8; DB 2; Length 110000;

```
  Best Local Similarity   49.1%;  Pred. No. 1.3e-14;
  Matches  370;  Conservative    0;  Mismatches  377;  Indels    7;  Gaps    2;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||| |     |  | ||   |||| |||| |||| |||| |          | |
Db      19974 ATGAATATGCCACTCATTAGTATAATAATGCCTGTTTATAATGCAGAATGTTACCTTAAT 20033

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
                |    |   |||     || ||||||||||||| |  |||  ||   | | |||||
Db      20034 CAAGGTATTTTATCTTGCTTAAATCAATCTTATCAAAACATTGAATTAATTTTAATTGAT 20093

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||| ||   || ||| | |   ||  | |||| || |          | ||  || ||
Db      20094 GATGGTTCAACAGATAAAAGTATTGAAATAATAAATAACATCATTGATAAAGATAAGCGG 20153

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              | | | | ||   || |   || |||  |||    |||||       |||||   ||  ||
Db      20154 GTTAAACTTTTCTTTACCCCAACTAATCAAGGGCCTGCAGCGGCAAGAAATATAGGGTTA 20213

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||  ||    ||  | |||||| | || |  |||   ||||||||        |  |
Db      20214 GAAAAAGCACAAGGTGACTATATTACATTCTTAGATAGTGATGATTTTATTGCAAATGAT 20273

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||    ||| |  |  |  |      |  |   | | |   ||      |   ||
Db      20274 AAATTAGAAAAACAACTTAATTTTATGTTGCAAAATCATCTTGTAATGACGCATGGCAAC 20333

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||  | |   |||    |  |||||    |       |||||   ||  | |    ||
Db      20334 TATGCTTTCTGTGATTTGGAAGGAAATCAGATAAAATTAGTTACGACTAGTAAAAAAATT 20393

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              |||||  ||   | |        ||| |  |    |   |||| | | |   | |
Db      20394 GATTATTTAACCTTATTACAAGGTAACCAATTTAAAATAATGACTGTACTTGTTGAAAGA 20453

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTA-----TTTGATGT 535
              |      |         ||          || |||   ||||| |||||     | | |
Db      20454 GAATCTATTAAATTACTTAGGTTTCCTAATATTAAACACGAAGATTACGCCTTCTTTTTG 20513

Qy        536 GGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATT 595
              |  ||   | ||  | | |||||    | |||| |       ||  ||  |    | |
Db      20514 GATTGTTTAAAAGAAGTTAAACAATCTATACTATATAGCCATCAGGCAAGTTCATTTGTA 20573

Qy        596 ACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGA 655
              | |  || | || | |     | ||    |||||| |||||  |||||  |||||||  |
Db      20574 AGAATCGGTAAAGTCAGCG--TTTCATCAAATAAATTTAAAAGTGCAATCTGGACATTTA 20631

Qy        656 GTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTAT 715
              ||| |||     ||       |  |      |   |||| || | ||||||||| |  | |
Db      20632 ATATCTATTTTAAAAGAGAGAAACTAGGTGTAGTAAAATCAATTTATTATTTTATCCTCT 20691

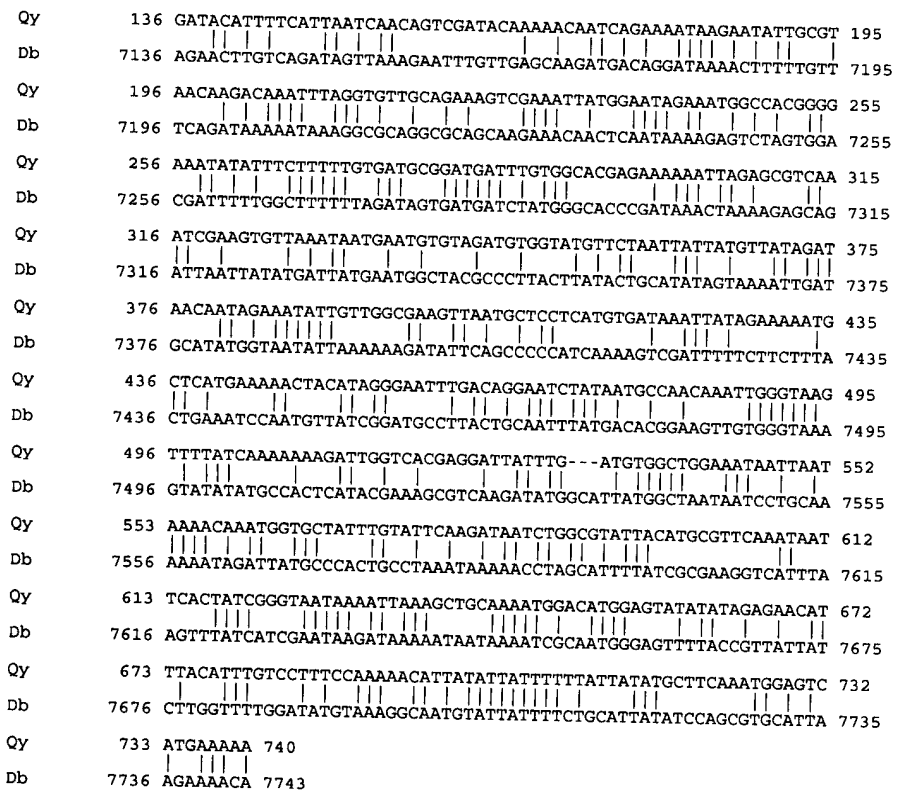
Qy        716 ATGCTTCAAATGGAGTCATGAAAAAAATAACACA 749
              |||| |  |||||| | || ||| | |  | | |
Db      20692 ATGCCTATAATGGATTTATTAAATACAAGAAATA 20725


RESULT 15
ADN02611
ID   ADN02611 standard; DNA; 9227 BP.
XX
AC   ADN02611;
XX
DT   17-JUN-2004  (first entry)
XX
DE   S. dysenteriae serotype 8 O-antigen biosynthetic gene cluster (partial).
XX
KW   O-antigen biosynthetic gene cluster; O-antigen gene cluster; serotype 8;
KW   lipopolysaccharide antigen; detection; gene; ds.
XX
OS   Shigella dysenteriae; serotype 8.
XX
FH   Key             Location/Qualifiers
FT   CDS             1097. .1912
FT                   /*tag=  a
FT                   /product= "Orf1 protein"
FT   CDS             1899. .3125
FT                   /*tag=  b
FT                   /product= "Wzy protein (orf2)"
FT   CDS             3180. .4349
FT                   /*tag=  c
FT                   /product= "Wzx protein (orf3)"
FT                   /note= "The CDS stop position given is that defined in
FT                   Claim 3, rather than that represented in the sequence
FT                   shown on pages 26-29 (4339). Decoding the protein from
FT                   the start codon given produces a 389 amino acid sequence
FT                   with a stop codon ending at position 4349, rather than
FT                   4339"
FT   CDS             4419. .5285
FT                   /*tag=  d
FT                   /product= "Orf4 protein"
FT   CDS             5282. .5968
FT                   /*tag=  e
FT                   /product= "Orf5 protein"
FT   CDS             5970. .7001
FT                   /*tag=  f
FT                   /product= "Gne protein (orf6)"
FT   CDS             7004. .7750
FT                   /*tag=  g
FT                   /product= "Orf7 protein"
XX
PN   CN1429833-A.
XX
PD   16-JUL-2003.
XX
PF   20-JAN-2003; 2003CN-00100539.
XX
PR   20-JAN-2003; 2003CN-00100539.
XX
PA   (UYNA-) UNIV NANKAI.
XX
PI   Wang L,  Guo H,  Feng L;
XX
DR   WPI; 2003-845848/79.
XX
PT   Nucleotide specific to O-antigen of shigella dysenteriae 8.
XX
PS   Claim 1; SEQ ID NO 1; 29pp; Chinese.
XX
CC   The invention relates to the O-antigen biosynthetic gene cluster of
CC   Shigella dysenteriae serotype 8 (ADN02611). The O-antigen of Shigella
CC   dysenteriae is a lipopolysaccharide antigen. The invention also relates
CC   to PCR primers for the amplification of the 6-phosphogluconate
CC   dehydrogenase (gnd) gene (ADN02612-ADN02613) or glycosyltransferase gene,
CC   a process for obtaining the O-antigen gene cluster, and a method for
CC   detecting Shigella dysenteriae serotype 8 in the environment and in the
CC   human body. The present sequence represents the specifically claimed
CC   partial sequence of the Shigella dysenteriae serotype 8 O-antigen
CC   biosynthetic gene cluster. Note: The features table gives CDS locations
CC   for 7 open reading frames (ORFs) which have mostly been obtained from the
CC   locations indicated in the sequence given on pages 26-29 of the
CC   specification, with the exception of the CDS stop position for wzx (orf4)
CC   - see note attached to this feature.
XX
SQ   Sequence 9227 BP; 2851 A; 1428 C; 1863 G; 3085 T; 0 U; 0 Other;

  Query Match             13.9%;  Score 108.8;  DB 11;  Length 9227;
  Best Local Similarity   47.8%;  Pred. No. 4.2e-13;
  Matches  348;  Conservative    0;  Mismatches  377;  Indels    3;  Gaps    1;

Qy         16 GTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCA 75
              || ||||| |||||||||  ||| |||   |     || || |    |    || ||||
Db       7016 GTGTCAATTATTATGCCCAGTTATAATTCTGAATTCACAATTAAGGAAAGTATAAAATCG 7075

Qy         76 ATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGAT 135
                | ||| | ||| | |||     |||  |  ||||  || | ||| || ||    || |||
Db       7076 GTAATAGAACAAACATATTCTAATTGGGAGTTGTTAATTACTGATGACTGCTCAACAGAT 7135
```



Search completed: May 10, 2006, 10:59:12
Job time : 370.515 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 11:00:02 ; Search time 2507.79 Seconds
(without alignments)
14608.170 Million cell updates/sec

Title: US-09-423-093B-2_COPY_79_861
Perfect score: 783
Sequence: 1 atgaacaaagaaaccgtttc.....gaaaggagactaaaaagtga 783

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

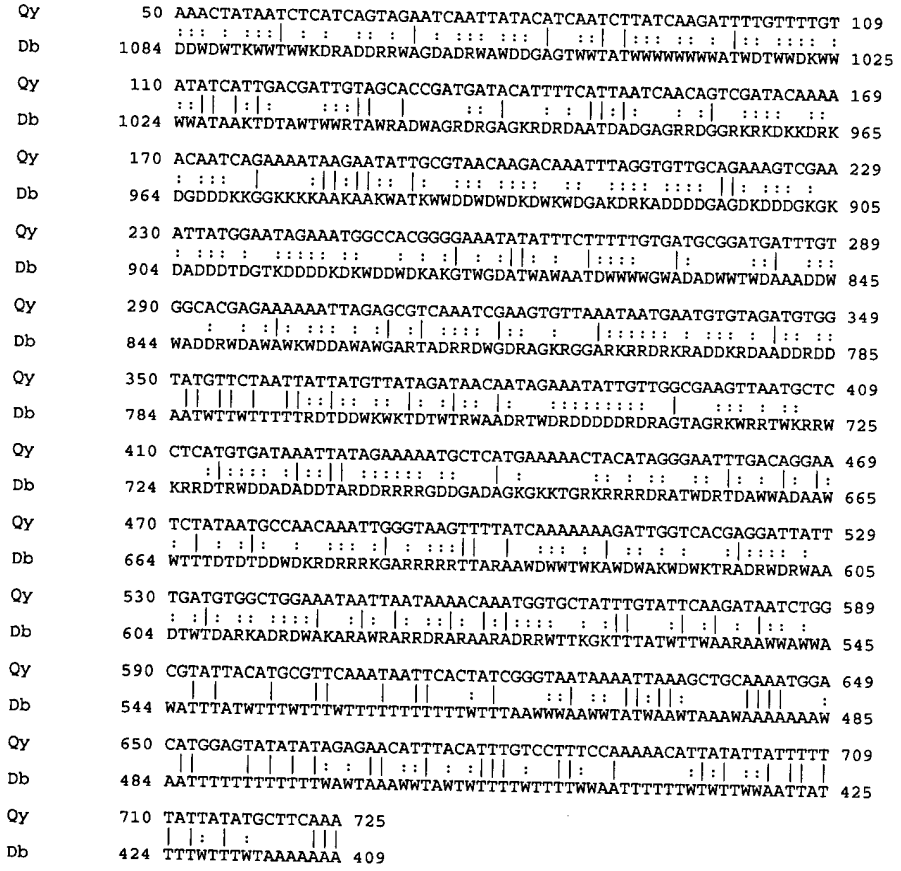
SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	80.4	10.3	1101	10	CNS0039G	AL063921 Drosophil
	2	72	9.2	1101	10	CNS00EVL	AL069706 Drosophil
	3	71.8	9.2	841	10	CZ541577	CZ541577 SRAA-aad3
	4	71.6	9.1	977	10	CNS0140Y	AL103564 Drosophil
	5	71.2	9.1	832	10	CZ541282	CZ541282 SRAA-aad3
c	6	65	8.3	1101	10	CNS00EVL	AL069706 Drosophil
c	7	64.8	8.3	1325	10	CW952344	CW952344 TcB38.1_D
c	8	64	8.2	879	10	CNS01JRG	AL147405 Anopheles
	9	64	8.2	1101	10	CNS003BD	AL064091 Drosophil
	10	63.8	8.1	959	10	CNS00655	AL062806 Drosophil
c	11	63.8	8.1	1101	10	CNS0039L	AL063926 Drosophil
c	12	63.6	8.1	891	10	CW952279	CW952279 TcB38.1_B
	13	63.4	8.1	1092	10	CNS020K7	AL175696 Tetraodon
c	14	63	8.0	720	10	CL677451	CL677451 PRI0120b_
	15	63	8.0	742	10	CL652986	CL652986 PRI0116b_
	16	63	8.0	773	10	CL673825	CL673825 PRI0110c_
c	17	63	8.0	829	10	CL683612	CL683612 PRI0137b_
	18	62.8	8.0	1101	10	CNS0006J	AL062049 Drosophil
c	19	62.4	8.0	1025	10	CNS014J2	AL104216 Drosophil
	20	62.2	7.9	1101	10	CNS008X3	AL052544 Drosophil
c	21	61.8	7.9	796	11	CT020317	CT020317 KBrH130P2
c	22	61.8	7.9	1101	10	CNS00ESI	AL069797 Drosophil
	23	61.8	7.9	1254	7	CF933916	CF933916 TrEST-B13
	24	61.6	7.9	807	10	CL696206	CL696206 PRI018b_G
	25	61.6	7.9	964	10	CNS006N9	AL065781 Drosophil
	26	61.4	7.8	475	1	AI724383	AI724383 RHIZ1_34_
	27	61.2	7.8	850	9	AZ670819	AZ670819 ENTHQ85TF
c	28	61	7.8	1392	10	CG757503	CG757503 P052-4-C0
	29	60.6	7.7	874	9	AZ536347	AZ536347 ENTBN59TR
	30	60.6	7.7	889	10	CW973509	CW973509 AIAA-aaa5
	31	60.4	7.7	1101	10	CNS00EO7	AL069440 Drosophil
	32	60.2	7.7	1101	10	CNS003B4	AL064082 Drosophil
c	33	60.2	7.7	1101	10	CNS00EO7	AL069440 Drosophil
	34	59.8	7.6	905	10	CNS00KHX	AL077798 Drosophil
	35	59.8	7.6	996	10	CNS00FUH	AL071063 Drosophil
	36	59.8	7.6	1148	1	AJ928602	AJ928602 AJ928602
	37	59.6	7.6	1086	10	CNS00YXK	AL096962 Drosophil
c	38	59.4	7.6	1587	10	AG347306	AG347306 Mus muscu
c	39	59.2	7.6	1113	7	CF994845	CF994845 AGENCOURT
	40	59.2	7.6	1117	1	AJ925762	AJ925762 AJ925762
c	41	59.2	7.6	1201	10	CNS0167M	AL106396 Drosophil
	42	59.2	7.6	1758	10	CL509408	CL509408 SAIL_811_
	43	59	7.5	1367	8	DN683496	DN683496 CGX30-C12
	44	58.8	7.5	876	9	AZ668287	AZ668287 ENTFW17TR
	45	58.8	7.5	987	10	CNS00KXT	AL078370 Drosophil

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

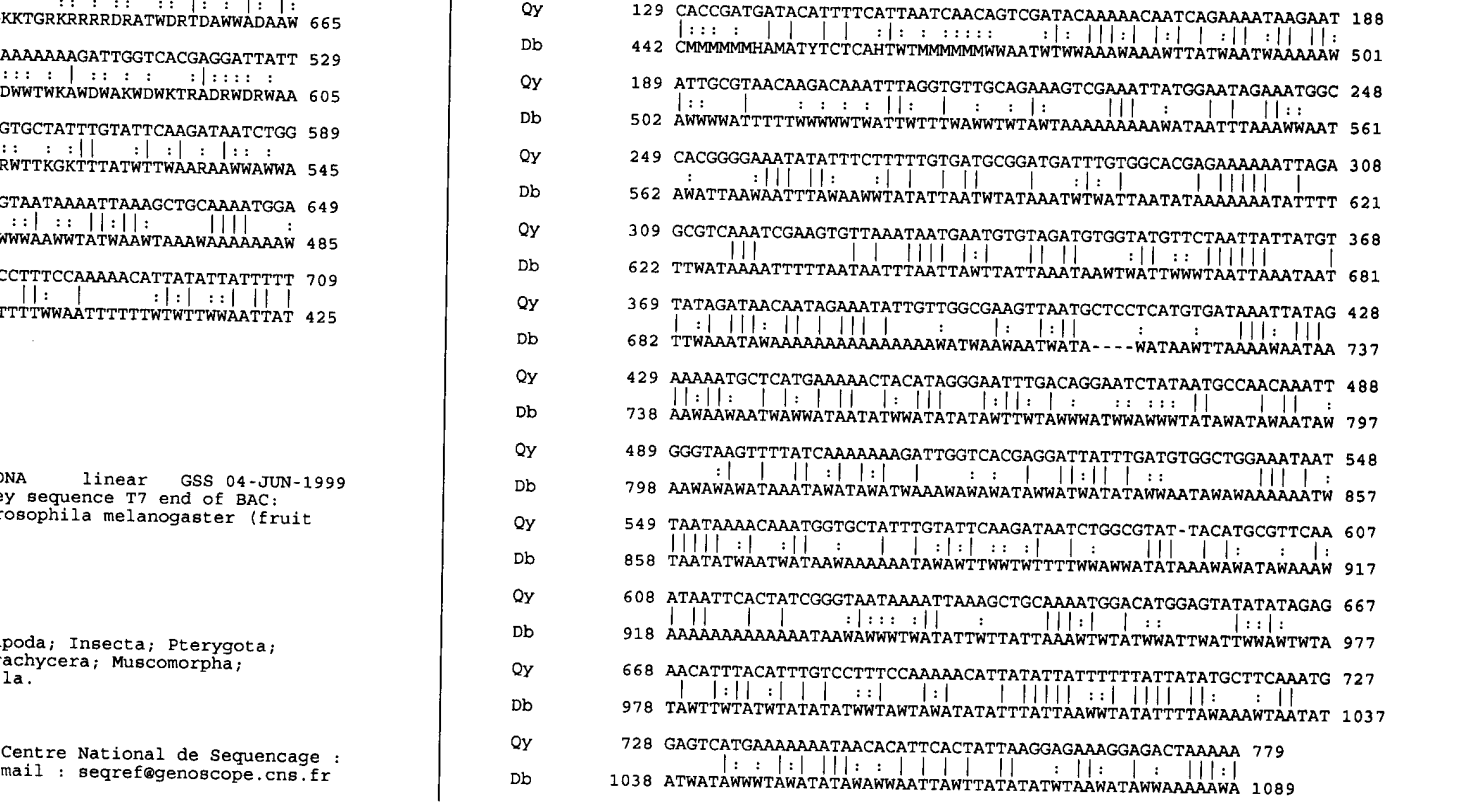
Query Match 10.3%; Score 80.4; DB 10; Length 1101;
Best Local Similarity 18.2%; Pred. No. 5.2e-06;
Matches 123; Conservative 278; Mismatches 275; Indels 0; Gaps 0;



RESULT 2
CNS00EVL
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"
ORIGIN

Query Match 9.2%; Score 72; DB 10; Length 1101;
Best Local Similarity 33.4%; Pred. No. 0.00025;
Matches 218; Conservative 114; Mismatches 315; Indels 5; Gaps 2;



```
RESULT 3
CZ541577
LOCUS       CZ541577               841 bp    DNA     linear   GSS 13-MAY-2005
DEFINITION  SRAA-aad39b05.g1 Strongyloides ratti whole genome shotgun library
            (SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.
ACCESSION   CZ541577
VERSION     CZ541577.1  GI:64667882
KEYWORDS    GSS.
SOURCE      Strongyloides ratti
  ORGANISM  Strongyloides ratti
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Panagrolaimoidea; Strongyloididae; Strongyloides.
REFERENCE   1  (bases 1 to 841)
  AUTHORS   Mitreva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,
            Ritter,E., Martin,J., Wylie,T., Dante,M., Waterston,R.H.,
            Clifton,S.W. and Wilson,R.
  TITLE     Genome Survey sequences from the rat parasitic nematode
            Strongyloides ratti
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Mitreva M
            Washington University in St. Louis
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: nematode@watson.wustl.edu
            Genomic DNA was provided by Fiona Thompson
            (F.Thompson@bristol.ac.uk) and Mark Viney
            (Mark.Viney@bristol.ac.uk) at the University of Bristol, Bristol,
            UK.
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .841
                     /organism="Strongyloides ratti"
                     /mol_type="genomic DNA"
                     /strain="Isofemale line ED321 heterogonic"
                     /db_xref="taxon:34506"
                     /dev_stage="infective larval stage (iL3)"
                     /lab_host="GS10"
                     /clone_lib="Strongyloides ratti whole genome shotgun
                     library (SRAAGSS 004)"
                     /note="Vector: pOTW13; Site_1: BstX1; Site_2: BstX1;
                     Strongyloides ratti genomic DNA was randomly sheared,
                     end-repaired and size fractioned to enrich for 2-4 kb
                     fragments. Genomic DNA was provided by Fiona Thompson
                     (F.Thompson@bristol.ac.uk) and Mark Viney
                     (Mark.Viney@bristol.ac.uk) at the University of Bristol,
                     Bristol, UK. Sequencing by Washington University Genome
                     Sequencing Center, St. Louis, MO."
ORIGIN

  Query Match             9.2%;  Score 71.8;  DB 10;  Length 841;
  Best Local Similarity   51.2%;  Pred. No. 0.00028;
  Matches  197;  Conservative    0;  Mismatches  182;  Indels    6;  Gaps    1;

Qy         16 GTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCA 75
              ||||| || ||||  |||   ||||||     |  ||  | |     | | || ||   ||
Db        209 GTTTCCATCATTACCCCCTGCTACAATTCTTCACAATTTCTGGAATAAACAATAACGTCT 268

Qy         76 ATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGAT 135
                |  || | ||| ||| |  |||||  |  | |   || | ||  ||  |    || ||
Db        269 GTACTAAACCAAACTTTTACAGATTGGGAATGGCTGATTACAGATGACAGATCTACAGAC 328

Qy        136 GATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGT 195
                || |  | | | | ||  | |       | | |||| ||| || ||||| | |      |
Db        329 AATTCTGTGTAAATGATTGAAA------AAATAAACGATCCGAGAATAAAACTTACTGTT 382

Qy        196 AACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGG 255
                | | |||   ||| |  | | |     |||||  |  | || |||  || || ||
Db        383 GCAGAAAAAAACGGAGGAGCAGGACATGCAAGAAATATTTCATTAAAAAATGCTACAGGA 442

Qy        256 AAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAA 315
```

```
              | || |||  | |||   ||||| |||||||| ||| |     ||    | |
Db        443 AGATTTATCACCTTTCTGGATGCAGATGATTTTTGGGAACCAGAATTTCTTCAAGAAATG 502

Qy        316 ATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGAT 375
              |  |   |  | || ||| ||   | |||  | |  |  |||| ||||          |||
Db        503 ACTGCTTTTATGAAAAATCAAAACGCAGAAATTGCCTACTCTACTTATGCAAGATGTGAT 562

Qy        376 AACAATAGAAATATTGTTGGCGAAG 400
               | |||   ||   | ||   ||||
Db        563 GAAAATTTCAAGCCTATTATTGAAG 587


RESULT 4
CNS0140Y
LOCUS       CNS0140Y                977 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN11P04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL103564
VERSION     AL103564.1  GI:5615175
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 977)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .977
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN11P04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             9.1%;  Score 71.6;  DB 10;  Length 977;
  Best Local Similarity   31.5%;  Pred. No. 0.00031;
  Matches  202;  Conservative   35;  Mismatches  403;  Indels    1;  Gaps    1;

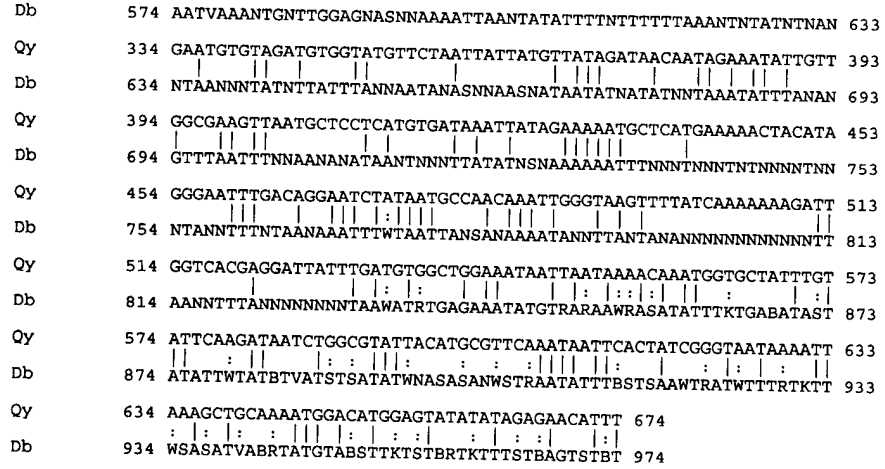
Qy         35 TTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAATTATACATCAATCTTATC 94
              |  | ||| : |  |   :              ||   ||  ||    ||  | ||
Db        334 TNGANAATTSTGNNATTNSNNNNNATAATNNNNNNNAAANTATNNTATNNTAANTTAATT 393

Qy         95 AAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATGATACATTTTCATTAATCA 154
                 ||       |    |  ||   ||| |                    |     |
Db        394 NNANNNTTNNAAATSAATNAANSGANATTTGAANNNNNATNNNNNANNNNNANTNNNNNA 453

Qy        155 ACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTAACAAGACAAATTTAGGTG 214
              | | |  | |:  |   | |       | |   || | |   || | | | |||||    |
Db        454 AAAWTANAAASTNATNNANTANTTTTANANANATTTATAATANCNAAAAAAATTNNNATN 513

Qy        215 TTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA-AATATATTTCTTTTTGT 273
               | |||     |   |    | ||      | |   | |||  |     || ||
Db        514 NNNNAAAAATANANTAAATTNTNNNAAAANNNANAATGANNANAATTTNAAATTTATTNA 573

Qy        274 GATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAATCGAAGTGTTAAATAAT 333
                 ||:    || || | |     | ||  || || |   |    |  ||    ||  | |
```



RESULT 5
CZ541282
LOCUS CZ541282 832 bp DNA linear GSS 13-MAY-2005
DEFINITION SRAA-aad37b05.g1 Strongyloides ratti whole genome shotgun library (SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.
ACCESSION CZ541282
VERSION CZ541282.1 GI:64667097
KEYWORDS GSS.
SOURCE Strongyloides ratti
ORGANISM Strongyloides ratti
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Panagrolaimoidea; Strongyloididae; Strongyloides.
REFERENCE 1 (bases 1 to 832)
AUTHORS Mitreva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D., Ritter,E., Martin,J., Wylie,T., Dante,M., Waterston,R.H., Clifton,S.W. and Wilson,R.
TITLE Genome Survey sequences from the rat parasitic nematode Strongyloides ratti
JOURNAL Unpublished (2005)
COMMENT Contact: Mitreva M
Washington University in St. Louis
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: nematode@watson.wustl.edu
Genomic DNA was provided by Fiona Thompson (F.Thompson@bristol.ac.uk) and Mark Viney (Mark.Viney@bristol.ac.uk) at the University of Bristol, Bristol, UK.
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .832
/organism="Strongyloides ratti"
/mol_type="genomic DNA"
/strain="Isofemale line ED321 heterogonic"
/db_xref="taxon:34506"
/dev_stage="infective larval stage (iL3)"
/lab_host="GS10"
/clone_lib="Strongyloides ratti whole genome shotgun library (SRAAGSS 004)"
/note="Vector: pOTW13; Site_1: BstX1; Site_2: BstX1; Strongyloides ratti genomic DNA was randomly sheared, end-repaired and size fractioned to enrich for 2-4 kb fragments. Genomic DNA was provided by Fiona Thompson (F.Thompson@bristol.ac.uk) and Mark Viney (Mark.Viney@bristol.ac.uk) at the University of Bristol, Bristol, UK. Sequencing by Washington University Genome Sequencing Center, St. Louis, MO."
ORIGIN

RESULT 6
CNS00EVL/c
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN
```

Query Match 8.3%; Score 65; DB 10; Length 1101;
Best Local Similarity 32.1%; Pred. No. 0.0065;
Matches 205; Conservative 109; Mismatches 324; Indels 1; Gaps 1;

```
Qy         88 TCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATGATACATTTTCA 147
Db       1100 TWWTATWTWTTTWTTTTTWWTATWTTAWATATATAAWTAATTWWTWTATATWTAWWWTAT 1041

Qy        148 TTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTAACAAGACAAAT 207
Db       1040 WATATATTAWTTTWTAAAATATAWWTTAATAAATATATWTAWTAWWATATATAWATAWAA 981

Qy        208 TTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGAAATATATTTCT 267
Db        980 WTATAWAWTWWAATWAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTT 921

Qy        268 TTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAATCGAAGTGTTA 327
Db        920 TTTWTTTWTATWTWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATA 861

Qy        328 AATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATAACAATAGAAAT 387
Db        860 TTAWATTTTTTWTWTATTWWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTWT 801

Qy        388 ATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGCTCATGAAAAAC 447
Db        800 WTTWTATTWTATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTT 741

Qy        448 TACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTGGGTAAGTTTTATCAAAAA 507
Db        740 WTTTTATTWTTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTWTATTTWAAA 681

Qy        508 AAGATTGGTCACGAGGATTATTTGATGTGGCTGGAAATAATTAATAAAACAAATGGT-GC 566
Db        680 TTATTTAATTAWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAA 621

Qy        567 TATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCAAATAATTCACTATCGGGTAA 626
Db        620 AAATATTTTTTTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTA 561

Qy        627 TAAAATTAAAGCTGCAAAATGGACATGGAGTATATATAGAGAACATTTACATTTGTCCTT 686
Db        560 TTWWTTTAAATTATWTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWTW 501

Qy        687 TCCAAAAACATTATATTATTTTTTATTATATGCTTCAAA 725
Db        500 TTTTTWATTWATAAWTTTWTTTWWAWATTWWKKKKKKAW 462
```

RESULT 7
CW952344/c

```
LOCUS       CW952344               1325 bp    DNA     linear   GSS 20-DEC-2004
DEFINITION  TcB38.1_D03_SP6 Tribolium BAC library Tribolium castaneum genomic,
            genomic survey sequence.
ACCESSION   CW952344
VERSION     CW952344.1  GI:56735365
KEYWORDS    GSS.
SOURCE      Tribolium castaneum (red flour beetle)
  ORGANISM  Tribolium castaneum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
            Tenebrionidae; Tribolium.
REFERENCE   1  (bases 1 to 1325)
  AUTHORS   Savard,J. and Tautz,D.
  TITLE     Tribolium castaneum BAC-ends sequencing project
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Savard, J.
            Abteilung fur Evolutionsgenetik, AG Tautz
            Institut fur Genetik, Universitat zu Koln
            Weyertal 121, 50931 Koln, Germany
            Tel: 49 221 470 6911
            Fax: 49 221 470 5975
            Email: savard@uni-koeln.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1325
                     /organism="Tribolium castaneum"
                     /mol_type="genomic DNA"
                     /strain="GA-2"
                     /db_xref="taxon:7070"
                     /clone_lib="Tribolium BAC library"
                     /note="Vector: pBACe3.6; Site_1: EcoR1; Site_2: EcoR1;
                     Library constructed by Exelixis Inc."
ORIGIN
```

Query Match 8.3%; Score 64.8; DB 10; Length 1325;
Best Local Similarity 47.1%; Pred. No. 0.0072;
Matches 329; Conservative 0; Mismatches 363; Indels 7; Gaps 4;

```
Qy         71 AATCAATTATACATCA-ATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGC 129
Db       1313 AATAAATTATTCATAATATCTCATATTAATATAATTTTATAATAATATATTTTATAATAT 1254

Qy        130 ACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATA 189
Db       1253 AATTATATTATATATTTAATATATAATAATATATATAATAATAATAATTAATTATCAAAA 1194

Qy        190 TTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCC 249
Db       1193 ATATATAATAATATAA--TAATATAATAATTAAAATATTTTTAATTAATATATAAAAATA 1136

Qy        250 ACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAG 309
Db       1135 TTATTTAAATATATTTAATATATATATTATTATTATATTATATAATAAAATAATATTATA 1076

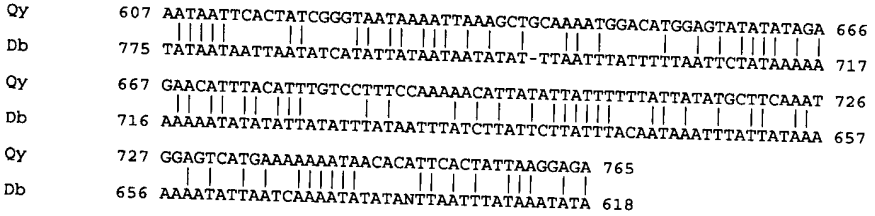
Qy        310 CGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTT 369
Db       1075 GAATAATACTAATGTAAAATATAAATATTTTATAATATACATAGTTAAATAATATTATTAT 1016

Qy        370 ATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGA 429
Db       1015 CTAATAAATTTAATAATTTTTTATACAAAAATAAATTCTTGTTATATTTATTTCATATAA 956

Qy        430 AAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTG 489
Db        955 AATATAAGTATATTAATATATAAATATAATAATTTATTTAATTATAATGATAAAAATATT 896

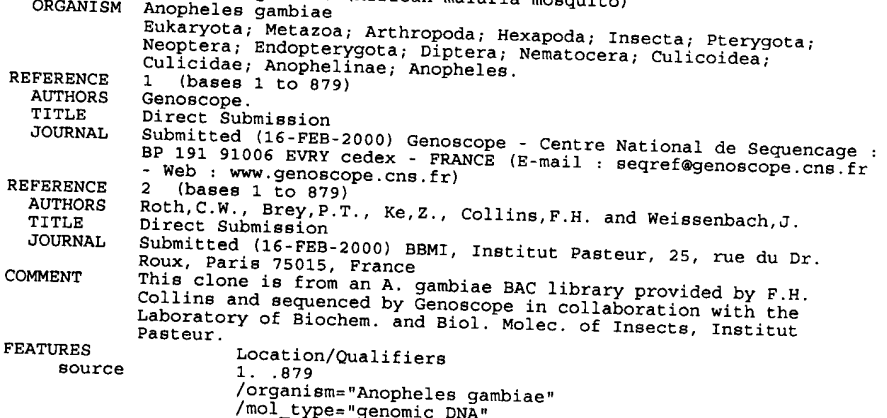
Qy        490 GGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGGAAATAATT 549
Db        895 TAATTTATCTATATTAATATTATTTATATAAATATTAATATTATTTTATTAATAATTATT 836

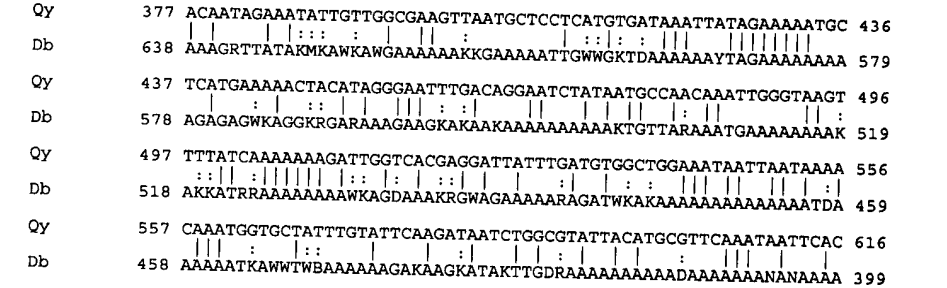
Qy        550 AATAA---AACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCA 606
Db        835 AATGATTTAACTAATTAATATATATATAATCAATATTATATTAAATATTAATTGAAATTA 776
```



RESULT 8
CNS01JRG/c
LOCUS CNS01JRG 879 bp DNA linear GSS 12-JUN-2001
DEFINITION Anopheles gambiae GSS T7 end of clone 14D07 of NotreDame1 library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.
ACCESSION AL147405
VERSION AL147405.1 GI:7005551
KEYWORDS GSS.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 879)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
REFERENCE 2 (bases 1 to 879)
AUTHORS Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France
COMMENT This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.
FEATURES Location/Qualifiers
source 1. .879
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="14D07"
/clone_lib="NotreDame1"
/note="end : T7"
ORIGIN

Query Match 8.2%; Score 64; DB 10; Length 879;
Best Local Similarity 38.7%; Pred. No. 0.01;
Matches 249; Conservative 52; Mismatches 341; Indels 1; Gaps 1;

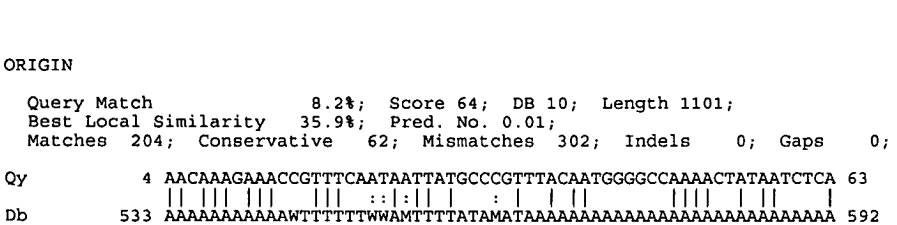




RESULT 9
CNS003BD
LOCUS CNS003BD 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL064091
VERSION AL064091.1 GI:4941847
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K08"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

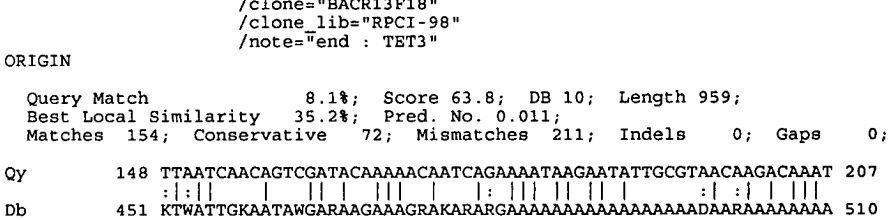
Query Match 8.2%; Score 64; DB 10; Length 1101;
Best Local Similarity 35.9%; Pred. No. 0.01;
Matches 204; Conservative 62; Mismatches 302; Indels 0; Gaps 0;



RESULT 10
CNS00655
LOCUS CNS00655 959 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR13F18 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL062806
VERSION AL062806.1 GI:4943996
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 959)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .959
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR13F18"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 8.1%; Score 63.8; DB 10; Length 959;
Best Local Similarity 35.2%; Pred. No. 0.011;
Matches 154; Conservative 72; Mismatches 211; Indels 0; Gaps 0;



RESULT 11
CNS0039L/c
LOCUS CNS0039L 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR08I10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063926
VERSION AL063926.1 GI:4941783
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission

```
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
  COMMENT   Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
  FEATURES             Location/Qualifiers
       source          1. .1101
                       /organism="Drosophila melanogaster"
                       /mol_type="genomic DNA"
                       /db_xref="taxon:7227"
                       /clone="BACR08I10"
                       /clone_lib="RPCI-98"
                       /note="end : T7"
  ORIGIN

    Query Match             8.1%;  Score 63.8;  DB 10;  Length 1101;
    Best Local Similarity   28.1%;  Pred. No. 0.011;
    Matches  137;  Conservative  131;  Mismatches  218;  Indels    1;  Gaps    1;

Qy        295 GAGAAAAAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGT 354
Db       1042 KWRWAWRAAWWDWGATGDTWTADWGDADDRRAADRRADRRRAADATARKRKKTGWTGARA 983

Qy        355 TCTAATTATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCAT 414
Db        982 TRTKTRTRWTTWRAWAAGTRATKWGDTDARTWTGTTTARTATGKKGTATWKGTTGKTTKK 923

Qy        415 GTGATAAATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTAT 474
Db        922 TTKRGGTKRKTKKAKAWAAAAAAAAAAAAAAAWAWWTTATATAATDTKWRATAGWAWATTK 863

Qy        475 AATGCCAACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATG 534
Db        862 ADATADWAWADATKRTKTATRAATTDKDAAAAAAAAAARATGAAAARARAAAAAARAKWTD 803

Qy        535 TGGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTAT 594
Db        802 DRWAAARWDAAARDARGTGGTAARDGWGGKKAAAKWTGRTTAAAADAWAAAAAAAATWAW 743

Qy        595 TACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGG 654
Db        742 WWAATAAKTAAAAAATARWGWAAAATKWTKAAAAAADAWAAARAGDKTAAAAAAAAARAR 683

Qy        655 AGTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTA 714
Db        682 AG-RGGKRTRKKRARAAATAAAATAAAAAAAAAAAADAWTAKKAAWAARTTTRTGGKRTGK 624

Qy        715 TATGCTTCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACT 774
Db        623 TKGTGGKGAAWWAGTATKRKAAAAAKTWKAAAAAAAAAAGKRTGGKKKKTRTBTRDWNB 564

Qy        775 AAAAAGT 781
Db        563 NAHAAAT 557


RESULT 12
CW952279/c
LOCUS       CW952279               891 bp    DNA     linear   GSS 20-DEC-2004
DEFINITION  TcB38.1_B02_T7 Tribolium BAC library Tribolium castaneum genomic,
            genomic survey sequence.
ACCESSION   CW952279
VERSION     CW952279.1  GI:56735316
KEYWORDS    GSS.
SOURCE      Tribolium castaneum (red flour beetle)
  ORGANISM  Tribolium castaneum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
            Tenebrionidae; Tribolium.
REFERENCE   1  (bases 1 to 891)
  AUTHORS   Savard,J. and Tautz,D.
  TITLE     Tribolium castaneum BAC-ends sequencing project
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Savard, J.
            Abteilung fur Evolutionsgenetik, AG Tautz
            Institut fur Genetik, Universitat zu Koln
            Weyertal 121, 50931 Koln, Germany
            Tel: 49 221 470 6911
            Fax: 49 221 470 5975
            Email: savard@uni-koeln.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .891
                     /organism="Tribolium castaneum"
                     /mol_type="genomic DNA"
                     /strain="GA-2"
                     /db_xref="taxon:7070"
                     /clone_lib="Tribolium BAC library"
                     /note="Vector: pBACe3.6; Site_1: EcoR1; Site_2: EcoR1;
                     Library constructed by Exelixis Inc."
ORIGIN

  Query Match             8.1%;  Score 63.6;  DB 10;  Length 891;
  Best Local Similarity   43.2%;  Pred. No. 0.013;
  Matches  303;  Conservative    0;  Mismatches  399;  Indels    0;  Gaps    0;

Qy         49 AAAACTATAATCTCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTG 108
Db        755 ATAATAATAATAATATAATAATATAATAATTATAATAATAATAATAATAATAATAATAAT 696

Qy        109 TATATCATTGACGATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAA 168
Db        695 AATAATAATAATAATGATAGTAATAATAATAATAATAATAATAATAATAGTAATAATAAT 636

Qy        169 AACAATCAGAAAATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGA 228
Db        635 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 576

Qy        229 AATTATGGAATAGAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTG 288
Db        575 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 516

Qy        289 TGGCACGAGAAAAAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTG 348
Db        515 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 456

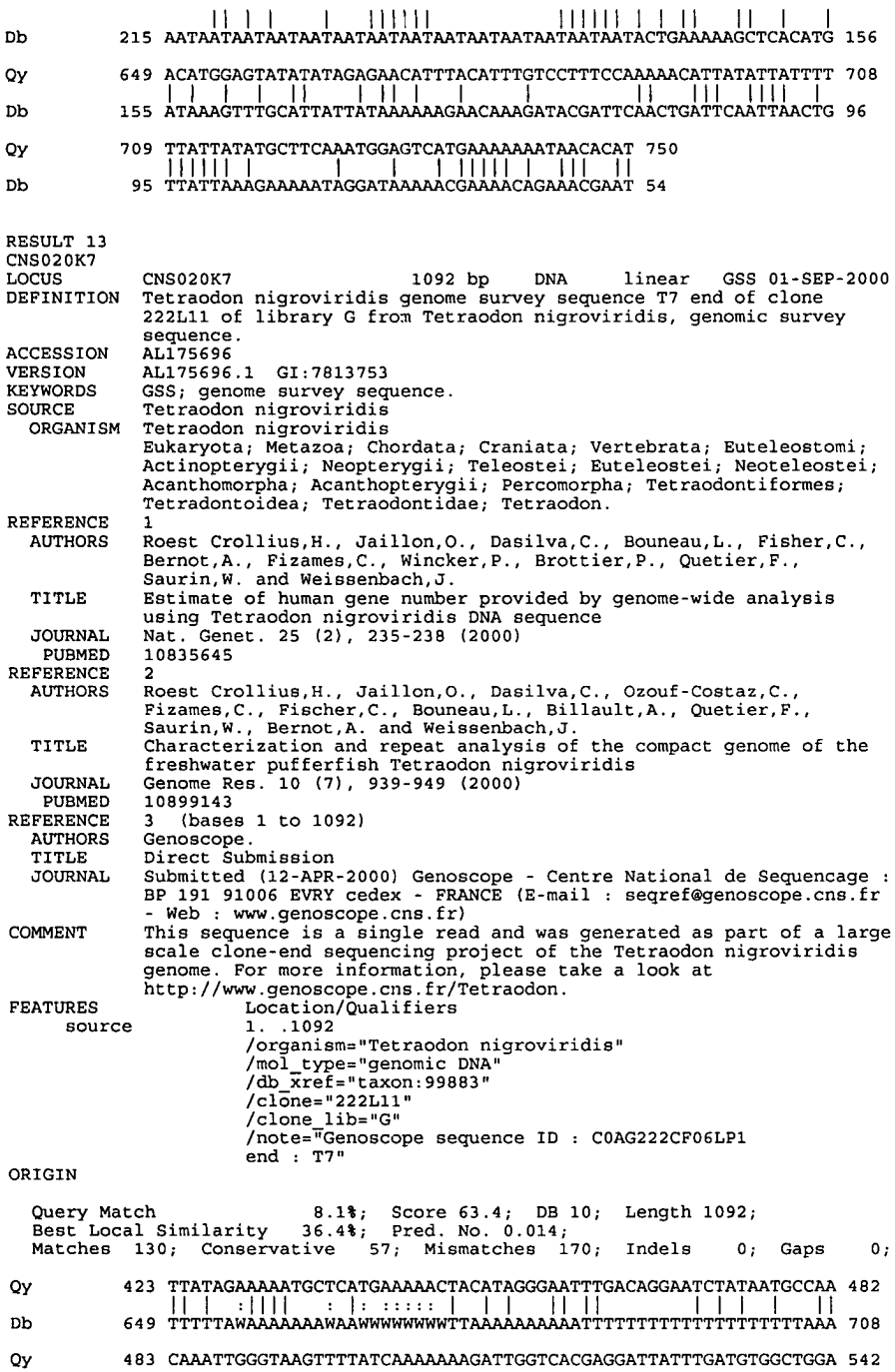
Qy        349 GTATGTTCTAATTATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCT 408
Db        455 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 396

Qy        409 CCTCATGTGATAAATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGA 468
Db        395 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 336

Qy        469 ATCTATAATGCCAACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTAT 528
Db        335 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 276

Qy        529 TTGATGTGGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTG 588
Db        275 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 216

Qy        589 GCGTATTACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGG 648
```



```
Db        215 AATAATAATAATAATAATAATAATAATAATAATAATAATAATACTGAAAAAGCTCACATG 156

Qy        649 ACATGGAGTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTT 708

Db        155 ATAAAGTTTGCATTATTATAAAAAAGAACAAAGATACGATTCAACTGATTCAATTAACTG 96

Qy        709 TTATTATATGCTTCAAATGGAGTCATGAAAAAAATAACACAT 750

Db         95 TTATTAAAGAAAAATAGGATAAAAACGAAAACAGAAACGAAT 54


RESULT 13
CNS020K7
LOCUS       CNS020K7                1092 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            222L11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL175696
VERSION     AL175696.1  GI:7813753
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1092)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="222L11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG222CF06LP1
                     end : T7"
ORIGIN

  Query Match             8.1%;  Score 63.4;  DB 10;  Length 1092;
  Best Local Similarity   36.4%;  Pred. No. 0.014;
  Matches  130;  Conservative   57;  Mismatches  170;  Indels    0;  Gaps    0;

Qy        423 TTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAA 482

Db        649 TTTTTAWAAAAAAAAWAAWWWWWWWTTAAAAAAAAAATTTTTTTTTTTTTTTTTTAAA 708

Qy        483 CAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGGA 542

Db        709 AAAAAAAAAAAAAAAWTWAATWAAAWWAAAAAAWAAWAAWWWAWTTWTTTTTTWAWWWWWW 768

Qy        543 AATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCG 602

Db        769 WAAAAWAAAWWWTWTTAWTTTTAWWWATWAWWTTWTATTTTAAAWAAAANAWTTTAWTTA 828

Qy        603 TTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATATA 662

Db        829 WTAAAAAAWWAAATAATTWTTTAAAAWAWTTTAWAAATTWWAATWTAAAAAAAAAAATTA 888

Qy        663 TAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCTTC 722

Db        889 AAAAAWAAWTWWTTTTTTTTWTWAWWTWTAAATWTTTTTTTTTTTTTTWWTTTTWTAAWA 948

Qy        723 AAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAA 779

Db        949 AAAAAAAAAAAATATAAAWNAAAAWWTTAATTTATTTTTAAAAAAAAAAAAAAATAA 1005


RESULT 14
CL677451/c
LOCUS       CL677451                720 bp    DNA     linear   GSS 09-JUL-2004
DEFINITION  PRI0120b_D10_2 - PRI0120b.BR (720) Mixed stage fosmid library of P.
            pacificus var. California Pristionchus pacificus genomic, genomic
            survey sequence.
ACCESSION   CL677451
VERSION     CL677451.1  GI:50183459
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
  TITLE     AppaDB: an AcedB database for the nematode satellite organism
            Pristionchus pacificus
  JOURNAL   Nucleic Acids Res. 32 (1), D421-D422 (2004)
   PUBMED   14681447
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            This library was generated at Caltech, Pasadena, USA and end
            sequenced at Vancouver, Canada.
            Seq primer: T7
            Class: fosmid ends.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Mixed stage fosmid library of P. pacificus
                     var. California"
                     /note="Vector: pEpifos-5 Fosmid vector"
ORIGIN

  Query Match             8.0%;  Score 63;  DB 10;  Length 720;
  Best Local Similarity   55.1%;  Pred. No. 0.017;
  Matches  145;  Conservative    0;  Mismatches  115;  Indels    3;  Gaps    1;

Qy         23 TAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAATTATAC 82

Db        484 TTATTATTCCGTTATATAATGCGGGCGATGATTTCCGCACTTGTATGGAATCTTTAATTA 425

Qy         83 ATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATGATACAT 142

Db        424 CGCAAACCTGGACTGCTCTGGAAATCATTATTATTAACGATGGTTCAACGGATAATTCTG 365
```

```
Qy        143 TTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTAACAAGA 202
              ||  | ||   ||    |        ||||| ||| | |  |  | | |||| |   ||
Db        364 TTGAAATAGCAAAGTATTACGCAGAAAACTATCCGCACGTTCGTTTGTTGCATC---AGG 308

Qy        203 CAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGAAATATA 262
              | |||   || |   | |      || |||  ||| || ||| |||| ||||| ||||||
Db        307 CGAATGCTGGCGCATCGGTGGCGCGTAATCGTGGGATTGAAGTGGCAACGGGCAAATATG 248

Qy        263 TTTCTTTTTGTGATGCGGATGAT 285
              |  |||||   |||||  || |||
Db        247 TCGCTTTTGTCGATGCTGACGAT 225


RESULT 15
CL652986
LOCUS       CL652986                742 bp    DNA     linear   GSS 09-JUL-2004
DEFINITION  PRI0116b_D06 - PRI0116b.B21 (742) Mixed stage fosmid library of P.
            pacificus var. California Pristionchus pacificus genomic, genomic
            survey sequence.
ACCESSION   CL652986
VERSION     CL652986.1  GI:50131764
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 742)
  AUTHORS   Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
  TITLE     AppaDB: an AcedB database for the nematode satellite organism
            Pristionchus pacificus
  JOURNAL   Nucleic Acids Res. 32 (1), D421-D422 (2004)
   PUBMED   14681447
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            This library was generated at Caltech, Pasadena, USA and end
            sequenced at Vancouver, Canada.
            Seq primer: T7
            Class: fosmid ends.
FEATURES             Location/Qualifiers
     source          1. .742
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Mixed stage fosmid library of P. pacificus
                     var. California"
                     /note="Vector: pEpifos-5 Fosmid vector"
ORIGIN

  Query Match             8.0%;  Score 63;  DB 10;  Length 742;
  Best Local Similarity   55.1%;  Pred. No. 0.017;
  Matches  145;  Conservative    0;  Mismatches  115;  Indels    3;  Gaps    1;

Qy         23 TAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAATTATAC 82
              | ||||| ||  | || |||| || | |   | |   | | |   | |||||  | ||
Db        170 TTATTATTCCGTTATATAATGCGGGCGATGATTTCCGCACTTGTATGGAATCTTTAATTA 229

Qy         83 ATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATGATACAT 142
                |||  | |      | | | |   |   ||| ||| ||||| ||   || ||| || |
Db        230 CGCAAACCTGGACTGCTCTGGAAATCATTATTATTAACGATGGTTCAACGGATAATTCTG 289

Qy        143 TTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTAACAAGA 202
              ||  | ||  ||  |        ||||| ||| | |  |  | | |||| |    ||
Db        290 TTGAAATAGCAAAGTATTACGCAGAAAACTATCCGCACGTTCGTTTGTTGCATC---AGG 346

Qy        203 CAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGAAATATA 262
              | |||   || |   | |      || |||  ||| || ||| |||| ||||| ||||||
Db        347 CGAATGCTGGCGCATCGGTGGCGCGTAATCGTGGGATTGAAGTGGCAACGGGCAAATATG 406

Qy        263 TTTCTTTTTGTGATGCGGATGAT 285
              |  |||||   |||||  || |||
Db        407 TCGCTTTTGTCGATGCTGACGAT 429


Search completed: May 11, 2006, 05:00:27
Job time : 2512.79 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 20:34:42 ; Search time 109.217 Seconds
(without alignments)
12743.667 Million cell updates/sec

Title: US-09-423-093B-2_COPY_79_861
Perfect score: 783
Sequence: 1 atgaacaaagaaaccgtttc.....gaaaggagactaaaaagtga 783

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	781.4	99.8	14187	3	US-09-453-702B-121	Sequence 121, App
c 2	781.4	99.8	14187	3	US-10-114-170-121	Sequence 121, App
3	116.8	14.9	1830121	3	US-09-557-884-1	Sequence 1, Appli
4	116.8	14.9	1830121	3	US-09-643-990A-1	Sequence 1, Appli
5	116.8	14.9	1830121	3	US-10-158-865-1	Sequence 1, Appli
6	100	12.8	891	3	US-09-816-028A-30	Sequence 30, Appl
7	100	12.8	891	3	US-10-303-162-30	Sequence 30, Appl
8	100	12.8	891	3	US-10-303-134-30	Sequence 30, Appl
9	100	12.8	891	3	US-10-303-118-30	Sequence 30, Appl
10	100	12.8	891	3	US-10-303-128-30	Sequence 30, Appl
11	93.2	11.9	906	3	US-09-495-406-14	Sequence 14, Appl
12	93.2	11.9	906	3	US-09-816-028A-26	Sequence 26, Appl
13	93.2	11.9	906	3	US-10-303-162-26	Sequence 26, Appl
14	93.2	11.9	906	3	US-10-303-134-26	Sequence 26, Appl
15	93.2	11.9	906	3	US-10-303-118-26	Sequence 26, Appl
16	93.2	11.9	906	3	US-10-303-128-26	Sequence 26, Appl
c 17	93.2	11.9	11474	3	US-09-495-406-1	Sequence 1, Appli
c 18	93.2	11.9	11474	3	US-09-816-028A-1	Sequence 1, Appli
c 19	93.2	11.9	11474	3	US-10-303-162-1	Sequence 1, Appli
c 20	93.2	11.9	11474	3	US-10-303-134-1	Sequence 1, Appli
c 21	93.2	11.9	11474	3	US-10-303-118-1	Sequence 1, Appli
c 22	93.2	11.9	11474	3	US-10-303-128-1	Sequence 1, Appli
c 23	85.2	10.9	1664976	3	US-08-916-421B-1	Sequence 1, Appli
c 24	85.2	10.9	1664976	3	US-09-692-570-1	Sequence 1, Appli
25	72.8	9.3	1002	3	US-09-107-532A-1469	Sequence 1469, Ap
26	68.2	8.7	939	3	US-09-900-038A-2	Sequence 2, Appli
27	68.2	8.7	6865	3	US-09-900-038A-3	Sequence 3, Appli
28	64.6	8.3	1170	3	US-09-495-406-11	Sequence 11, Appl
29	64.6	8.3	1170	3	US-09-816-028A-15	Sequence 15, Appl
30	64.6	8.3	1170	3	US-10-303-162-15	Sequence 15, Appl
31	64.6	8.3	1170	3	US-10-303-134-15	Sequence 15, Appl
32	64.6	8.3	1170	3	US-10-303-118-15	Sequence 15, Appl
33	64.6	8.3	1170	3	US-10-303-128-15	Sequence 15, Appl
34	64.6	8.3	11474	3	US-09-495-406-1	Sequence 1, Appli
35	64.6	8.3	11474	3	US-09-816-028A-1	Sequence 1, Appli
36	64.6	8.3	11474	3	US-10-303-162-1	Sequence 1, Appli
37	64.6	8.3	11474	3	US-10-303-134-1	Sequence 1, Appli
38	64.6	8.3	11474	3	US-10-303-118-1	Sequence 1, Appli
39	64.6	8.3	11474	3	US-10-303-128-1	Sequence 1, Appli
40	63.2	8.1	14602	2	US-08-597-236-1	Sequence 1, Appli
41	63.2	8.1	14602	2	US-08-746-682A-1	Sequence 1, Appli
42	62.2	7.9	912	3	US-09-495-406-16	Sequence 16, Appl
43	62.2	7.9	912	3	US-09-816-028A-28	Sequence 28, Appl
44	62.2	7.9	912	3	US-10-303-162-28	Sequence 28, Appl
45	62.2	7.9	912	3	US-10-303-134-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-09-453-702B-121/c
; Sequence 121, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14187
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-09-453-702B-121

```
  Query Match             99.8%;  Score 781.4;  DB 3;  Length 14187;
  Best Local Similarity   99.9%;  Pred. No. 1.8e-169;
  Matches  782;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13922 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 13863

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13862 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 13803

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13802 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 13743

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13742 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 13683

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13682 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 13623

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13622 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 13563

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      13562 TATTATGTTATAGATAACAAAAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 13503

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13502 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 13443

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13442 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 13383

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13382 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 13323

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13322 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 13263

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13262 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 13203

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13202 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 13143

Qy        781 TGA 783
              |||
Db      13142 TGA 13140


RESULT 2
US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Patent No. 6855814
;   GENERAL INFORMATION:
;       APPLICANT: Blattner, Frederick R.
;                  Burland,    Valerie
;                  Perna,  Nicole T.
;                  Plunkett, Guy
;                  Welch,  Rod
;       TITLE OF INVENTION: No. 6855814el Sequences of E. coli O157
;       NUMBER OF SEQUENCES: 265
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Quarles & Brady
;            STREET: 1 South Pinckney Street
;            CITY: Madison
;            STATE: WI
;            COUNTRY: US
;            ZIP: 53701-2113
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: Word Perfect 8.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/114,170
;            FILING DATE: 01-Apr-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/453,702
;            FILING DATE: 03-DEC-1999
;            APPLICATION NUMBER: 60/110,955
;            FILING DATE: 04-DEC-1998
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Seay, Nicholas J.
;            REGISTRATION NUMBER: 27386
;            REFERENCE/DOCKET NUMBER: 960296.95017
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (608) 251-5000
;            TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 121:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 14187
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA  (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121

  Query Match             99.8%;  Score 781.4;  DB 3;  Length 14187;
  Best Local Similarity   99.9%;  Pred. No. 1.8e-169;
  Matches  782;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13922 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 13863

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13862 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 13803

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13802 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 13743

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13742 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 13683

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13682 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 13623

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13622 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 13563

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      13562 TATTATGTTATAGATAACAAAAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 13503

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      13502 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 13443

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13442 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 13383

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13382 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 13323

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13322 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 13263

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13262 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 13203

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13202 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 13143

Qy        781 TGA 783
              |||
Db      13142 TGA 13140


RESULT 3
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
;   GENERAL INFORMATION:
;        APPLICANT: Fleischmann et al.
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/557,884
;             FILING DATE: 25-Apr-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/476,102
;             FILING DATE: JUN-5-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Michelle S. Marks
;             REGISTRATION NUMBER: 41,971
;             REFERENCE/DOCKET NUMBER: PB186P3
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-309-8504
;             TELEFAX: 301-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

  Query Match             14.9%;  Score 116.8;  DB 3;  Length 1830121;
  Best Local Similarity   49.1%;  Pred. No. 5.9e-17;
  Matches  370;  Conservative    0;  Mismatches  377;  Indels    7;  Gaps    2;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||| |     |  | ||    ||||| ||||| ||||| ||||  |          | |
Db     919974 ATGAATATGCCACTCATTAGTATAATAATGCCTGTTTATAATGCAGAATGTTACCTTAAT 920033

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
                |   |||      || ||||||||||||||  |  |||  ||   | | |||||
Db     920034 CAAGGTATTTTATCTTGCTTAAATCAATCTTATCAAAACATTGAATTAATTTTAATTGAT 920093

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||| ||   || ||| | |   ||  | |||| || |          | ||  || ||
Db     920094 GATGGTTCAACAGATAAAAGTATTGAAATAATAAATAACATCATTGATAAAGATAAGCGG 920153

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              | | | | ||   || |   || ||| |||   |||||       |||||   ||  ||
Db     920154 GTTAAACTTTTCTTTACCCCAACTAATCAAGGGCCTGCAGCGGCAAGAAATATAGGGTTA 920213

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||  ||    ||  | |||||| | || |  |||   |||||||       | |
Db     920214 GAAAAAGCACAAGGTGACTATATTACATTCTTAGATAGTGATGATTTTATTGCAAATGAT 920273

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||    ||| |  |  |  |      |  |    | | |   ||      |   ||
Db     920274 AAATTAGAAAAACAACTTAATTTTATGTTGCAAAATCATCTTGTAATGACGCATGGCAAC 920333

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              ||| | |    |||     | |||||    |       |||||   ||  | |    ||
Db     920334 TATGCTTTCTGTGATTTGGAAGGAAATCAGATAAAATTAGTTACGACTAGTAAAAAAATT 920393

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              |||||  ||   |  |         ||| |  |     |   |||| | | |   | |
Db     920394 GATTATTTAACCTTATTACAAGGTAACCAATTTAAAATAATGACTGTACTTGTTGAAAGA 920453

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTA-----TTTGATGT 535
              |     |         ||          || |||   ||||| |||||     | |  |
Db     920454 GAATCTATTAAATTACTTAGGTTTCCTAATATTAAACACGAAGATTACGCCTTCTTTTTG 920513

Qy        536 GGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATT 595
              |  ||   | ||  | | ||||||    | |||| |        ||  ||  |    | |
Db     920514 GATTGTTTAAAAGAAGTTAAACAATCTATACTATATAGCCATCAGGCAAGTTCATTTGTA 920573

Qy        596 ACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGA 655
              | |  || | || | |     | ||    |||||| ||||| |||||  |||||||  |
Db     920574 AGAATCGGTAAAGTCAGCG--TTTCATCAAATAAATTTAAAAGTGCAATCTGGACATTTA 920631

Qy        656 GTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTAT 715
              ||| |||    ||        |  |      |  |||| || | |||||||||  |  | |
Db     920632 ATATCTATTTTAAAAGAGAGAAACTAGGTGTAGTAAAATCAATTTATTATTTTATCCTCT 920691

Qy        716 ATGCTTCAAATGGAGTCATGAAAAAAATAACACA 749
              |||| |  |||||| | || ||| | |  | | |
Db     920692 ATGCCTATAATGGATTTATTAAATACAAGAAATA 920725


RESULT 4
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
;   GENERAL INFORMATION:
;        APPLICANT: Robert D. Fleischmann
;                   Mark D. Adams
;                   Owen White
;                   Hamilton O. Smith
;                   J. Craig Venter
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
```

```
;             CITY: Rockville,
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;       COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/643,990A
;             FILING DATE: 23-Aug-2000
;             CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/487,429
;             FILING DATE: 1995-06-07
;             APPLICATION NUMBER: 08/426,787
;             FILING DATE: 1995-04-21
;       ATTORNEY/AGENT INFORMATION:
;             NAME: Kenley K. Hoover
;             REGISTRATION NUMBER: 40,302
;             REFERENCE/DOCKET NUMBER: PB186P1C1
;       TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-610-5790
;             TELEFAX: 310-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

  Query Match           14.9%;  Score 116.8;  DB 3;  Length 1830121;
  Best Local Similarity  49.1%;  Pred. No. 5.9e-17;
  Matches  370;  Conservative    0;  Mismatches  377;  Indels    7;  Gaps    2;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||| |     |  | ||   ||||| ||||| ||||  |          | |
Db     919974 ATGAATATGCCACTCATTAGTATAATAATGCCTGTTTATAATGCAGAATGTTACCTTAAT 920033

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
                |   |  |||     || ||||||||||||||  |  |||  ||   | | |||||
Db     920034 CAAGGTATTTTATCTTGCTTAAATCAATCTTATCAAAACATTGAATTAATTTTAATTGAT 920093

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||| ||   || ||| | |   ||  | |||| || |          | ||  || ||
Db     920094 GATGGTTCAACAGATAAAAGTATTGAAATAATAAATAACATCATTGATAAAGATAAGCGG 920153

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              | | | | ||   || |   || |||  |||   |||||          |||||  ||  ||
Db     920154 GTTAAACTTTTCTTTACCCCAACTAATCAAGGGCCTGCAGCGGCAAGAAATATAGGGTTA 920213

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              |||| ||    ||  | ||||||  | || |  |||   ||||||||        |  |
Db     920214 GAAAAAGCACAAGGTGACTATATTACATTCTTAGATAGTGATGATTTTATTGCAAATGAT 920273

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||    ||| |  |  | |     |  |  | | |   ||     |   ||
Db     920274 AAATTAGAAAAACAACTTAATTTTATGTTGCAAAATCATCTTGTAATGACGCATGGCAAC 920333

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              ||| | |   |||    |  |||||    |       |||||   ||  | |    ||
Db     920334 TATGCTTTCTGTGATTTGGAAGGAAATCAGATAAAATTAGTTACGACTAGTAAAAAAATT 920393

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              |||||  ||   | |         ||| | |    |   |||| | | |     | |
Db     920394 GATTATTTAACCTTATTACAAGGTAACCAATTTAAAATAATGACTGTACTTGTTGAAAGA 920453

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTA-----TTTGATGT 535
              |      |          ||          || |||   ||||| |||||     | | |
Db     920454 GAATCTATTAAATTACTTAGGTTTCCTAATATTAAACACGAAGATTACGCCTTCTTTTTG 920513

Qy        536 GGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATT 595
              |  ||   | || | | ||||||    | |||| |       ||  ||  |    | |
Db     920514 GATTGTTTAAAAGAAGTTAAACAATCTATACTATATAGCCATCAGGCAAGTTCATTTGTA 920573

Qy        596 ACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGA 655
              | |  || | || | |     | ||    ||||| |||||  |||||  |||||| |
Db     920574 AGAATCGGTAAAGTCAGCG--TTTCATCAAATAAATTTAAAAGTGCAATCTGGACATTTA 920631

Qy        656 GTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTAT 715
              |||  |||    ||       | |      |   |||| || | |||||||||  |  | |
Db     920632 ATATCTATTTTAAAAGAGAGAAACTAGGTGTAGTAAAATCAATTTATTATTTTATCCTCT 920691

Qy        716 ATGCTTCAAATGGAGTCATGAAAAAAATAACACA 749
              |||| |  |||||| | || ||| | |  | | |
Db     920692 ATGCCTATAATGGATTTATTAAATACAAGAAATA 920725


RESULT 5
US-10-158-865-1
; Sequence 1, Application US/10158865
; Patent No. 6846651
; GENERAL INFORMATION:
;  APPLICANT: Fleischmann et al.
;  TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Fragm
; Patent No. 6846651
;  TITLE OF INVENTION:  Thereof, and Uses Thereof
;  FILE REFERENCE: PB186P2C1D1
;  CURRENT APPLICATION NUMBER: US/10/158,865
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US 09/557,884
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: US 08/476,102
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: US 08/426,787
;  PRIOR FILING DATE: 1995-04-21
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1830121
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4747)..(4747)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9921)..(9921)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (10150)..(10150)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (29298)..(29298)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36543)..(36543)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36551)..(36551)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36636)..(36636)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (40808)..(40810)
```

; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a,t,c, or g

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:

; NAME/KEY: misc_feature

Query Match 14.9%; Score 116.8; DB 3; Length 1830121;
Best Local Similarity 49.1%; Pred. No. 5.9e-17;
Matches 370; Conservative 0; Mismatches 377; Indels 7; Gaps 2;

```
Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
Db     919974 ATGAATATGCCACTCATTAGTATAATAATGCCTGTTTATAATGCAGAATGTTACCTTAAT 920033

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
Db     920034 CAAGGTATTTTATCTTGCTTAAATCAATCTTATCAAAACATTGAATTAATTTTAATTGAT 920093

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
Db     920094 GATGGTTCAACAGATAAAAGTATTGAAATAATAAATAACATCATTGATAAAGATAAGCGG 920153

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
Db     920154 GTTAAACTTTTCTTTACCCCAACTAATCAAGGGCCTGCAGCGGCAAGAAATATAGGGTTA 920213

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
Db     920214 GAAAAAGCACAAGGTGACTATATTACATTCTTAGATAGTGATGATTTTATTGCAAATGAT 920273

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
Db     920274 AAATTAGAAAAACAACTTAATTTTATGTTGCAAAATCATCTTGTAATGACGCATGGCAAC 920333

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
Db     920334 TATGCTTTCTGTGATTTGGAAGGAAATCAGATAAAATTAGTTACGACTAGTAAAAAAATT 920393

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
Db     920394 GATTATTTAACCTTATTACAAGGTAACCAATTTAAAATAATGACTGTACTTGTTGAAAGA 920453

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTA-----TTTGATGT 535
Db     920454 GAATCTATTAAATTACTTAGGTTTCCTAATATTAAACACGAAGATTACGCCTTCTTTTTG 920513

Qy        536 GGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATT 595
Db     920514 GATTGTTTAAAAGAAGTTAAACAATCTATACTATATAGCCATCAGGCAAGTTCATTTGTA 920573

Qy        596 ACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGA 655
Db     920574 AGAATCGGTAAAGTCAGCG--TTTCATCAAATAAATTTAAAAGTGCAATCTGGACATTTA 920631

Qy        656 GTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTAT 715
Db     920632 ATATCTATTTTAAAAGAGAGAAACTAGGTGTAGTAAAATCAATTTATTATTTTATCCTCT 920691

Qy        716 ATGCTTCAAATGGAGTCATGAAAAAAATAACACA 749
Db     920692 ATGCCTATAATGGATTTATTAAATACAAGAAATA 920725
```

RESULT 6
US-09-816-028A-30
; Sequence 30, Application US/09816028A
; Patent No. 6699705
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(891)
; OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
US-09-816-028A-30

Query Match 12.8%; Score 100; DB 3; Length 891;
Best Local Similarity 48.5%; Pred. No. 9.3e-14;
Matches 333; Conservative 0; Mismatches 349; Indels 4; Gaps 2;

```
Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76
Db         11 TTTCAATCATCTTGCCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGTT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136
Db         71 GTATCAATCAGACTTTTAAAAATATAGAAATAATTGTAGTTGATGATTGTGGAAGTGACA 130

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
Db        131 AAAGTATAGATATAGTTAAAGAATATGCCAAAAAAGATGATAGAATAAAAATCATACATA 190

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256
Db        191 ATGAAGAAAATTTAAAACTTTTAAGAGCTAGATATGAAGGTGTAAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316
Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTAGAACTTAATGCTTGTGAAGAATGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376
Db        311 TGAAAATTTTAAAAAACAATGAAATAGATTTATTATTTTTTAATGCATTTGTATTGGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436
Db        371 ATAACAATAAAATAGAAAGAAAGTTGAA---TTTTCAAGAAAAATGTTATGTAAAAAAAG 427

Qy        437 TCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTGGGTAAGT 496
Db        428 ATTTTTTAAAAGAACTATTAAAAACTAAAAATTTATTTTGGACAGTGTGGGCAAAAGTCA 487

Qy        497 TTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGGAAATAATTAATAAAA 556
Db        488 TAAAAAAAGAATTATATCTCAAGGCTGTTGGTTTAATATCGCTAGAAAATGCTAAAATAA 547

Qy        557 CAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCAAATAATTCAC 616
Db        548 ATATGGCTGAAGATGTTTTATTATATTACCCTTTGATAAATATTTCAAATACTATATTTC 607

Qy        617 TATCGGGTAATAAAAT-TAAAGCTGCAAAATGGACATGGAGTATATATAGAGAACATTTA 675
Db        608 ACTTGAGTAAAAATTTATACAATTATCAAATAAATAATTTCTCTATAACCAAAACATTAA 667

Qy        676 CATTTGTCCTTTCCAAAAACATTATA 701
Db        668 CATTGCAAAATATAAAAACAAATATA 693
```

RESULT 7
US-10-303-162-30
; Sequence 30, Application US/10303162
; Patent No. 6723545
; GENERAL INFORMATION:

```
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,162
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
;   LENGTH: 891
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(891)
;   OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
US-10-303-162-30

  Query Match             12.8%;  Score 100;  DB 3;  Length 891;
  Best Local Similarity   48.5%;  Pred. No. 9.3e-14;
  Matches  333;  Conservative    0;  Mismatches  349;  Indels    4;  Gaps    2;

Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76
              ||||||| ||  ||||   ||| |||| ||   ||  ||||       || |||||
Db         11 TTTCAATCATCTTGCCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGTT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136
              |||  ||||  ||| | || || | |    |   | |  |||| |||||| | |  ||
Db         71 GTATCAATCAGACTTTTAAAAATATAGAAATAATTGTAGTTGATGATTGTGGAAGTGACA 130

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
              | |   |      || | ||          ||||||  || | | ||||| |||  | | ||
Db        131 AAAGTATAGATATAGTTAAAGAATATGCCAAAAAAGATGATAGAATAAAAATCATACATA 190

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256
              | |   |||||||    ||  |  | | || || | ||  || || | || |
Db        191 ATGAAGAAAATTTAAAACTTTTAAGAGCTAGATATGAAGGTGTAAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316
              |||||    ||||  ||| | |||||||  |   |    ||    |  ||       |
Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTAGAACTTAATGCTTGTGAAGAATGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376
              |  || | ||||| ||  |      ||||| |  ||| || ||||   | |||  | || |
Db        311 TGAAAATTTTAAAAAACAATGAAATAGATTTATTATTTTTTAATGCATTTGTATTGGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436
              | || |  || || |   |  |  | ||   |  ||| |  | |  ||||  |||||
Db        371 ATAACAATAAAATAGAAAGAAAGTTGAA---TTTTCAAGAAAAATGTTATGTAAAAAAAG 427

Qy        437 TCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTGGGTAAGT 496
                 |   |||  |  ||   ||     | |    | | |     |
Db        428 ATTTTTTAAAAGAACTATTAAAAACTAAAAATTTATTTTGGACAGTGTGGGCAAAAGTCA 487

Qy        497 TTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGGAAATAATTAATAAAA 556
              |  |  || ||     |   | |  | |  ||| || | ||| ||||    ||| | ||
Db        488 TAAAAAAAGAATTATATCTCAAGGCTGTTGGTTTAATATCGCTAGAAAATGCTAAAATAA 547

Qy        557 CAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCAAATAATTCAC 616
              |  | ||    |  | | | || | |    |    | ||          | ||  |  |
Db        548 ATATGGCTGAAGATGTTTTATTATATTACCCTTTGATAAATATTTCAAATACTATATTTC 607

Qy        617 TATCGGGTAATAAAAT-TAAAGCTGCAAAATGGACATGGAGTATATATAGAGAACATTTA 675
              | | |||| ||  | || |  |   ||||  | |     | ||||   | ||||||  |
Db        608 ACTTGAGTAAAAATTTATACAATTATCAAATAAATAATTTCTCTATAACCAAAACATTAA 667

Qy        676 CATTTGTCCTTTCCAAAAACATTATA 701
              ||||      |   ||||  | ||||
Db        668 CATTGCAAAATATAAAAACAAATATA 693


RESULT 8
US-10-303-134-30
; Sequence 30, Application US/10303134
; Patent No. 6825019
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,134
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
;   LENGTH: 891
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(891)
;   OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
US-10-303-134-30

  Query Match             12.8%;  Score 100;  DB 3;  Length 891;
  Best Local Similarity   48.5%;  Pred. No. 9.3e-14;
  Matches  333;  Conservative    0;  Mismatches  349;  Indels    4;  Gaps    2;

Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76
              ||||||| ||  ||||   ||| |||| ||   ||  ||||       || |||||
Db         11 TTTCAATCATCTTGCCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGTT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136
              |||  ||||  ||| | || || | |    |   | |  |||| |||||| | |  ||
Db         71 GTATCAATCAGACTTTTAAAAATATAGAAATAATTGTAGTTGATGATTGTGGAAGTGACA 130

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
              | |   |      || | ||          ||||||  || | | ||||| |||  | | ||
Db        131 AAAGTATAGATATAGTTAAAGAATATGCCAAAAAAGATGATAGAATAAAAATCATACATA 190

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256
              | |   |||||||    ||  |  | | || || | ||  || || | || |
Db        191 ATGAAGAAAATTTAAAACTTTTAAGAGCTAGATATGAAGGTGTAAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316
              |||||    ||||  ||| | |||||||  |   |    ||    |  ||       |
Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTAGAACTTAATGCTTGTGAAGAATGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376
              |  || | ||||| ||  |      ||||| |  ||| || ||||   | |||  | || |
Db        311 TGAAAATTTTAAAAAACAATGAAATAGATTTATTATTTTTTAATGCATTTGTATTGGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436
              | || |  || || |   |  |  | ||   |  ||| |  | |  ||||  |||||
Db        371 ATAACAATAAAATAGAAAGAAAGTTGAA---TTTTCAAGAAAAATGTTATGTAAAAAAAG 427

Qy        437 TCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTGGGTAAGT 496
                 |   |||  |  ||   ||     | |    | | |     |
Db        428 ATTTTTTAAAAGAACTATTAAAAACTAAAAATTTATTTTGGACAGTGTGGGCAAAAGTCA 487
```

```
Qy        497 TTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGGAAATAATTAATAAAA 556
Db        488 TAAAAAAAGAATTATATCTCAAGGCTGTTGGTTTAATATCGCTAGAAAATGCTAAAATAA 547

Qy        557 CAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCAAATAATTCAC 616
Db        548 ATATGGCTGAAGATGTTTTATTATATTACCCTTTGATAAATATTTCAAATACTATATTTC 607

Qy        617 TATCGGGTAATAAAAT-TAAAGCTGCAAAATGGACATGGAGTATATATAGAGAACATTTA 675
Db        608 ACTTGAGTAAAAATTTATACAATTATCAAATAAATAATTTCTCTATAACCAAAACATTAA 667

Qy        676 CATTTGTCCTTTCCAAAAACATTATA 701
Db        668 CATTGCAAAATATAAAAACAAATATA 693
```

RESULT 9
US-10-303-118-30
; Sequence 30, Application US/10303118
; Patent No. 6905867
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,118
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(891)
; OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
US-10-303-118-30

Query Match 12.8%; Score 100; DB 3; Length 891;
Best Local Similarity 48.5%; Pred. No. 9.3e-14;
Matches 333; Conservative 0; Mismatches 349; Indels 4; Gaps 2;

```
Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76
Db         11 TTTCAATCATCTTGCCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGTT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136
Db         71 GTATCAATCAGACTTTTAAAAATATAGAAATAATTGTAGTTGATGATTGTGGAAGTGACA 130

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
Db        131 AAAGTATAGATATAGTTAAAGAATATGCCAAAAAAGATGATAGAATAAAAATCATACATA 190

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256
Db        191 ATGAAGAAAATTTAAAACTTTTAAGAGCTAGATATGAAGGTGTAAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316
Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTAGAACTTAATGCTTGTGAAGAATGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376
Db        311 TGAAAATTTTAAAAAACAATGAAATAGATTTATTATTTTTTAATGCATTTGTATTGGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436
Db        371 ATAACAATAAAATAGAAAGAAAGTTGAA---TTTTCAAGAAAAATGTTATGTAAAAAAAG 427

Qy        437 TCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTGGGTAAGT 496
Db        428 ATTTTTTAAAAGAACTATTAAAAACTAAAAATTTATTTTGGACAGTGTGGGCAAAAGTCA 487

Qy        497 TTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGGAAATAATTAATAAAA 556
Db        488 TAAAAAAAGAATTATATCTCAAGGCTGTTGGTTTAATATCGCTAGAAAATGCTAAAATAA 547

Qy        557 CAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCAAATAATTCAC 616
Db        548 ATATGGCTGAAGATGTTTTATTATATTACCCTTTGATAAATATTTCAAATACTATATTTC 607

Qy        617 TATCGGGTAATAAAAT-TAAAGCTGCAAAATGGACATGGAGTATATATAGAGAACATTTA 675
Db        608 ACTTGAGTAAAAATTTATACAATTATCAAATAAATAATTTCTCTATAACCAAAACATTAA 667

Qy        676 CATTTGTCCTTTCCAAAAACATTATA 701
Db        668 CATTGCAAAATATAAAAACAAATATA 693
```

RESULT 10
US-10-303-128-30
; Sequence 30, Application US/10303128
; Patent No. 6911337
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,128
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(891)
; OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
US-10-303-128-30

Query Match 12.8%; Score 100; DB 3; Length 891;
Best Local Similarity 48.5%; Pred. No. 9.3e-14;
Matches 333; Conservative 0; Mismatches 349; Indels 4; Gaps 2;

```
Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76
Db         11 TTTCAATCATCTTGCCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGTT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136
Db         71 GTATCAATCAGACTTTTAAAAATATAGAAATAATTGTAGTTGATGATTGTGGAAGTGACA 130

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
```

```
Db        131 AAAGTATAGATATAGTTAAAGAATATGCCAAAAAAGATGATAGAATAAAAATCATACATA 190

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256

Db        191 ATGAAGAAAATTTAAAACTTTTAAGAGCTAGATATGAAGGTGTAAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316

Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTAGAACTTAATGCTTGTGAAGAATGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376

Db        311 TGAAAATTTTAAAAAACAATGAAATAGATTTATTATTTTTTAATGCATTTGTATTGGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436

Db        371 ATAACAATAAAATAGAAAGAAAGTTGAA---TTTTCAAGAAAAATGTTATGTAAAAAAAG 427

Qy        437 TCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTGGGTAAGT 496

Db        428 ATTTTTTAAAAGAACTATTAAAAACTAAAAATTTATTTTGGACAGTGTGGGCAAAAGTCA 487

Qy        497 TTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGGAAATAATTAATAAAA 556

Db        488 TAAAAAAAGAATTATATCTCAAGGCTGTTGGTTTAATATCGCTAGAAAATGCTAAAATAA 547

Qy        557 CAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCAAATAATTCAC 616

Db        548 ATATGGCTGAAGATGTTTTATTATATTACCCTTTGATAAATATTTCAAATACTATATTTC 607

Qy        617 TATCGGGTAATAAAAT-TAAAGCTGCAAAATGGACATGGAGTATATATAGAGAACATTTA 675

Db        608 ACTTGAGTAAAAATTTATACAATTATCAAATAAATAATTTCTCTATAACCAAAACATTAA 667

Qy        676 CATTTGTCCTTTCCAAAAACATTATA 701

Db        668 CATTGCAAAATATAAAAACAAATATA 693
```

RESULT 11
US-09-495-406-14
; Sequence 14, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000110US
; CURRENT APPLICATION NUMBER: US/09/495,406
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(906)
; OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
; OTHER INFORMATION: OH4384 (ORF 6a of LOS biosynthesis locus)
US-09-495-406-14

Query Match 11.9%; Score 93.2; DB 3; Length 906;
Best Local Similarity 50.7%; Pred. No. 3.4e-12;
Matches 224; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

```
Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76

Db         11 TTTCAATCATCTTACCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGCT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136

Db         71 GTATCAATCAGACTTTTAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATA 130

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196

Db        131 ATAGTATAAATATAGCCAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACA 190

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256

Db        191 ATGAAAAAAACTTAGGTCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316

Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376

Db        311 TAAAAATTTTAGATGAACAGGATGAAGTTGATTTAGTGTTTTTCAATGCTATTGTTGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436

Db        371 GTAATGTTATTTCATATAAAAAGTTTGACTTTAATTCTGGTTTTTATAGCAAAAAAGAGT 430

Qy        437 TCATGAAAAACTACATAGGGAA 458

Db        431 TTGTAAAAAAAATTATTGCAAA 452
```

RESULT 12
US-09-816-028A-26
; Sequence 26, Application US/09816028A
; Patent No. 6699705
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(906)
; OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
; OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
; OTHER INFORMATION: biosynthesis locus)
US-09-816-028A-26

Query Match 11.9%; Score 93.2; DB 3; Length 906;
Best Local Similarity 50.7%; Pred. No. 3.4e-12;
Matches 224; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

```
Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76

Db         11 TTTCAATCATCTTACCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGCT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136

Db         71 GTATCAATCAGACTTTTAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATA 130
```

```
Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
Db        131 ATAGTATAAATATAGCCAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACA 190

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256
Db        191 ATGAAAAAAACTTAGGTCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316
Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376
Db        311 TAAAAATTTTAGATGAACAGGATGAAGTTGATTTAGTGTTTTTCAATGCTATTGTTGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436
Db        371 GTAATGTTATTTCATATAAAAAGTTTGACTTTAATTCTGGTTTTTATAGCAAAAAAGAGT 430

Qy        437 TCATGAAAAACTACATAGGGAA 458
Db        431 TTGTAAAAAAAATTATTGCAAA 452
```

RESULT 13
US-10-303-162-26
; Sequence 26, Application US/10303162
; Patent No. 6723545
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,162
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(906)
; OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
; OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
; OTHER INFORMATION: biosynthesis locus)
US-10-303-162-26

Query Match 11.9%; Score 93.2; DB 3; Length 906;
Best Local Similarity 50.7%; Pred. No. 3.4e-12;
Matches 224; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

```
Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76
Db         11 TTTCAATCATCTTACCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGCT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136
Db         71 GTATCAATCAGACTTTTAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATA 130

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
Db        131 ATAGTATAAATATAGCCAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACA 190

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256
Db        191 ATGAAAAAAACTTAGGTCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316
Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376
Db        311 TAAAAATTTTAGATGAACAGGATGAAGTTGATTTAGTGTTTTTCAATGCTATTGTTGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436
Db        371 GTAATGTTATTTCATATAAAAAGTTTGACTTTAATTCTGGTTTTTATAGCAAAAAAGAGT 430

Qy        437 TCATGAAAAACTACATAGGGAA 458
Db        431 TTGTAAAAAAAATTATTGCAAA 452
```

RESULT 14
US-10-303-134-26
; Sequence 26, Application US/10303134
; Patent No. 6825019
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,134
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(906)
; OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
; OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
; OTHER INFORMATION: biosynthesis locus)
US-10-303-134-26

Query Match 11.9%; Score 93.2; DB 3; Length 906;
Best Local Similarity 50.7%; Pred. No. 3.4e-12;
Matches 224; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

```
Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76
Db         11 TTTCAATCATCTTACCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGCT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136
Db         71 GTATCAATCAGACTTTTAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATA 130

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
Db        131 ATAGTATAAATATAGCCAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACA 190

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256
```

```
                |  | | ||| |||||| ||  |  |    || || | ||  |  || | || |
Db        191 ATGAAAAAAACTTAGGTCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316
                |||||    ||||  ||| | |||||||  | | |    ||     |  ||        |
Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376
                |  || | ||| || |  |    || || ||   ||  | |    ||  | || | || |
Db        311 TAAAAATTTTAGATGAACAGGATGAAGTTGATTTAGTGTTTTTCAATGCTATTGTTGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436
                  |||    | |      |    |  || |   |  | ||   |  ||   | ||||  |
Db        371 GTAATGTTATTTCATATAAAAAGTTTGACTTTAATTCTGGTTTTTATAGCAAAAAAGAGT 430

Qy        437 TCATGAAAAACTACATAGGGAA 458
                |  | |||||    || |  ||
Db        431 TTGTAAAAAAAAATTATTGCAAA 452


RESULT 15
US-10-303-118-26
; Sequence 26, Application US/10303118
; Patent No. 6905867
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,118
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-303-118-26

  Query Match             11.9%;  Score 93.2;  DB 3;  Length 906;
  Best Local Similarity   50.7%;  Pred. No. 3.4e-12;
  Matches  224;  Conservative    0;  Mismatches  218;  Indels    0;  Gaps    0;

Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76
              ||||||| ||  | ||   ||| |||| ||   ||  ||||        || |||||
Db         11 TTTCAATCATCTTACCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGCT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136
               |||  ||||  ||| | ||||| | |    |   | |  |||| ||||||  | |  |||
Db         71 GTATCAATCAGACTTTTAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATA 130

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
              |||    |      ||  |||        |  |||||  |  | | ||||| |||| | |  |
Db        131 ATAGTATAAATATAGCCAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACA 190

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256
              |  | | ||| |||||| ||  |  |    || || | ||  |  || | || |
Db        191 ATGAAAAAAACTTAGGTCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316
                |||||    ||||  ||| | |||||||  | | |    ||     |  ||        |
Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376
                |  || | ||| || |  |    || || ||   ||  | |    ||  | || | || |
Db        311 TAAAAATTTTAGATGAACAGGATGAAGTTGATTTAGTGTTTTTCAATGCTATTGTTGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436
                  |||    | |      |    |  || |   |  | ||   |  ||   | ||||  |
Db        371 GTAATGTTATTTCATATAAAAAGTTTGACTTTAATTCTGGTTTTTATAGCAAAAAAGAGT 430

Qy        437 TCATGAAAAACTACATAGGGAA 458
                |  | |||||    || |  ||
Db        431 TTGTAAAAAAAAATTATTGCAAA 452


Search completed: May 11, 2006, 05:22:45
Job time : 118.217 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:01:11 ; Search time 516.803 Seconds
(without alignments)
12528.801 Million cell updates/sec

Title: US-09-423-093B-2_COPY_79_861
Perfect score: 783
Sequence: 1 atgaacaaagaaaccgtttc.....gaaaggagactaaaaagtga 783

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	783	100.0	14024	3	US-09-423-093A-2	Sequence 2, Appli
c 2	781.4	99.8	14187	5	US-10-114-170-121	Sequence 121, App
3	181.8	23.2	747	7	US-10-282-122A-17153	Sequence 17153, A
4	162.8	20.8	774	7	US-10-282-122A-16844	Sequence 16844, A
c 5	140.4	17.9	4915	3	US-09-070-927A-125	Sequence 125, App
6	116.8	14.9	753	7	US-10-282-122A-22150	Sequence 22150, A
7	116.8	14.9	1830121	7	US-10-329-670-1	Sequence 1, Appli
8	116.8	14.9	1830121	8	US-10-158-865-1	Sequence 1, Appli
9	116.8	14.9	1830121	9	US-10-981-687-1	Sequence 1, Appli
10	100	12.8	891	3	US-09-816-028A-30	Sequence 30, Appl
11	100	12.8	891	6	US-10-303-161-30	Sequence 30, Appl
12	100	12.8	891	6	US-10-303-118-30	Sequence 30, Appl
13	100	12.8	891	6	US-10-303-128-30	Sequence 30, Appl
14	100	12.8	891	6	US-10-303-134-30	Sequence 30, Appl
15	100	12.8	891	6	US-10-303-162-30	Sequence 30, Appl
16	100	12.8	891	8	US-10-735-419-30	Sequence 30, Appl
17	100	12.8	891	8	US-10-820-536-30	Sequence 30, Appl
18	100	12.8	891	8	US-10-845-408-30	Sequence 30, Appl
19	100	12.8	891	8	US-10-845-412-30	Sequence 30, Appl
20	100	12.8	891	8	US-10-846-219-30	Sequence 30, Appl
21	100	12.8	891	8	US-10-821-604-30	Sequence 30, Appl
22	100	12.8	891	8	US-10-847-983-30	Sequence 30, Appl
23	100	12.8	891	8	US-10-821-573-30	Sequence 30, Appl
24	100	12.8	891	8	US-10-850-807-30	Sequence 30, Appl
25	100	12.8	891	8	US-10-850-125-30	Sequence 30, Appl
26	100	12.8	891	8	US-10-830-825-30	Sequence 30, Appl
27	100	12.8	891	9	US-10-962-334-30	Sequence 30, Appl
28	100	12.8	891	9	US-10-830-997-30	Sequence 30, Appl
29	100	12.8	891	9	US-10-962-235-30	Sequence 30, Appl
30	100	12.8	891	9	US-10-961-882-30	Sequence 30, Appl
31	95.2	12.2	3738	3	US-09-767-041-43	Sequence 43, Appl
32	93.2	11.9	906	3	US-09-816-028A-26	Sequence 26, Appl
33	93.2	11.9	906	6	US-10-303-161-26	Sequence 26, Appl
34	93.2	11.9	906	6	US-10-303-118-26	Sequence 26, Appl
35	93.2	11.9	906	6	US-10-303-128-26	Sequence 26, Appl
36	93.2	11.9	906	6	US-10-303-134-26	Sequence 26, Appl
37	93.2	11.9	906	6	US-10-303-162-26	Sequence 26, Appl
38	93.2	11.9	906	8	US-10-735-419-26	Sequence 26, Appl
39	93.2	11.9	906	8	US-10-820-536-26	Sequence 26, Appl
40	93.2	11.9	906	8	US-10-845-408-26	Sequence 26, Appl
41	93.2	11.9	906	8	US-10-845-412-26	Sequence 26, Appl
42	93.2	11.9	906	8	US-10-846-219-26	Sequence 26, Appl
43	93.2	11.9	906	8	US-10-821-604-26	Sequence 26, Appl
44	93.2	11.9	906	8	US-10-847-983-26	Sequence 26, Appl
45	93.2	11.9	906	8	US-10-821-573-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-09-423-093A-2
; Sequence 2, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 14024
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-423-093A-2

Query Match 100.0%; Score 783; DB 3; Length 14024;
Best Local Similarity 100.0%; Pred. No. 1.5e-149;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 138

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 198

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 258

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 318

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 378
```

```
Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 438

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 498

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 558

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 618

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 678

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        679 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 738

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 798

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        799 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 858

Qy        781 TGA 783
              |||
Db        859 TGA 861
```

```
RESULT 2
US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Publication No. US20030023075A1
;   GENERAL INFORMATION:
;        APPLICANT: Blattner, Frederick R.
;                   Burland,     Valerie
;                   Perna,  Nicole T.
;                   Plunkett, Guy
;                   Welch,  Rod
;        TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
;        NUMBER OF SEQUENCES: 265
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Quarles & Brady
;             STREET: 1 South Pinckney Street
;             CITY: Madison
;             STATE: WI
;             COUNTRY: US
;             ZIP: 53701-2113
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Word Perfect 8.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/114,170
;             FILING DATE: 01-Apr-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/453,702
;             FILING DATE: 03-DEC-1999
;             APPLICATION NUMBER: 60/110,955
;             FILING DATE: 04-DEC-1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seay, Nicholas J.
;             REGISTRATION NUMBER: 27386
;             REFERENCE/DOCKET NUMBER: 960296.95017
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (608) 251-5000
;             TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 121:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14187
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA  (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121

  Query Match             99.8%;  Score 781.4;  DB 5;  Length 14187;
  Best Local Similarity   99.9%;  Pred. No. 3.2e-149;
  Matches  782;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13922 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 13863

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13862 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 13803

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13802 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 13743

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13742 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 13683

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13682 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 13623

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13622 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 13563

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      13562 TATTATGTTATAGATAACAAAAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 13503

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13502 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 13443

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13442 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTG 13383

Qy        541 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13382 GAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATG 13323

Qy        601 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13322 CGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATA 13263

Qy        661 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13262 TATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCT 13203

Qy        721 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13202 TCAAATGGAGTCATGAAAAAAATAACACATTCACTATTAAGGAGAAAGGAGACTAAAAAG 13143

Qy        781 TGA 783
              |||
```

```
Db      13142 TGA 13140


RESULT 3
US-10-282-122A-17153
; Sequence 17153, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17153
;   LENGTH: 747
;   TYPE: DNA
;   ORGANISM: Clostridium difficile
US-10-282-122A-17153

  Query Match             23.2%;  Score 181.8;  DB 7;  Length 747;
  Best Local Similarity   54.6%;  Pred. No. 3.2e-27;
  Matches  407;  Conservative    0;  Mismatches  332;  Indels    6;  Gaps    2;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||| ||  |    || || |||||||  ||| | || |||   |  |||  |||
Db          1 ATGAAGAAGAATTTAGTATCTATAATTACTCCCATGTATAATTCTGAAAAATTTATTGAA 60

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
              || |  || ||||| |  || | ||| ||||||||||  |  |   |||  || |||||
Db         61 GCAACCATAAAATCAGTATTAAACCAAACTTATCAAGAATGGGAAATGTTAATTATTGAT 120

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              |||||    || ||| |||    |    || ||||  |        || |  ||   | |
Db        121 GATTGCTCAACAGATAATAGTCCTAATATAGTCAAATCTTATATGCAACAGGATAGTAGA 180

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              |||| |    |    |   |||| |||   |||||  |  |   | ||||||  | | ||
Db        181 ATAAAATGTATAAAGACTGAGACTAATAAGGGTGTCTCTAATGCTAGAAATTTAGCACTA 240

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              | ||| || ||  ||| |||  ||||||  |||   |||||    ||| |    |
Db        241 AGTAAGGCAACAGGACAATTTATAGCTTTTTTAGATAGTGATGACCAATGGAATAGTAGT 300

Qy        301 AAATTAGAGCGTCAAATCGA---AGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCT 357
              || |||||    ||| |  |     ||||| | ||||| | ||||  |    |   |||
Db        301 AAGTTAGAAAAACAAGTAAATTTTATGTTAGAAAATGACTATGTAATTTCATTTACTTCA 360

Qy        358 AATTATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTG 417
              || |    |   || ||  ||| |||  |       | | | || | ||| | |
Db        361 TATGAACTGATGGATGAAAATGATAAAAAATTAAACAAAGTAATAAAAGTTCCACCTAAT 420

Qy        418 ATAAATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAAT 477
              || |||||||||| | |  | |    |||   | ||||   | | || |    |   | ||
Db        421 GTAGATTATAGAAGATTATTAAAAGGAAATATCTTAGGATGTCTTACTGTTGTGATTGAT 480

Qy        478 GCCAACAAATTGGGTAAGTTTTATCAAAAAAAGATTG---GTCACGAGGATTATTTGATG 534
                |||||||  |          ||  |    ||   | || || |||||| | |
Db        481 AAGTCTAAATTGGATTTTGAAATAAGAATGAGTGGTGTAAGACATGAAGATTATGTTTTA 540

Qy        535 TGGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTAT 594
              ||| |    ||| | || |||    ||  ||||  | | ||  | ||   ||| || |
Db        541 TGGTTATCGATATTGAAAAAAGGGCATATTGCTCATGGAATAAATGAAGTTCTAGCATTA 600

Qy        595 TACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGG 654
              || |    || |  ||||| |||   |||||||||| |||||||  |  ||||| |||
Db        601 TATAGAAAATCTAGCAATTCCCTAAGTGGTAATAAAATAAAAGCTGCTATGTGGACTTGG 660

Qy        655 AGTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTA 714
              | |||||||||| |      | |  |  | ||    ||| |  |||| || ||| |
Db        661 AATATATATAGAAATATAGAAAAAATACCTTTATACAAAGCTATATACTACTTTATTAAT 720

Qy        715 TATGCTTCAAATGGAGTCATGAAAA 739
              ||||    ||| ||| |||  ||||
Db        721 TATGGAATAAACGGAATCAAAAAAA 745


RESULT 4
US-10-282-122A-16844
; Sequence 16844, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
```

; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16844
; LENGTH: 774
; TYPE: DNA
; ORGANISM: Clostridium difficile
US-10-282-122A-16844

Query Match 20.8%; Score 162.8; DB 7; Length 774;
Best Local Similarity 52.9%; Pred. No. 2.5e-23;
Matches 397; Conservative 0; Mismatches 347; Indels 6; Gaps 2;

```
Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
Db         25 ATGAATGAGCCTTTAGTATCGATAATTACTCCTGTTTATAATTCAGAAGAGTTTTTATCA 84

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
Db         85 GAGACGATAAAGTCAATCCAAAATCAAACTTATAAAAATTGGCAGCTTTTATTAGTGGAT 144

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
Db        145 GATTGTTCTAAGGACAATTCAAGTGAAATTATAAAAAGTTTTCGAAAAGAAGATGCTAGA 204

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
Db        205 ATAAAATATATAAAGCTTGAAAAAAATTCTGGAGCTGCAGTATCTAGAAATGTTGGTATT 264

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
Db        265 AAGAATGCGGAAGGTCGATTTATTGCATTTGTAGATAGTGATGATTTATGGGATAGTAGA 324

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGT---TCT 357
Db        325 AAATTAGAGATTCAGATAGAATATATGTTAAAAGAAAATGTAGGTTTTTCTTTTACATCT 384

Qy        358 AATTATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTG 417
Db        385 TATAGATATATGAGACAAGATGGGTCAAAAACGAATAAGGTTGCAAGAGCTCCTAAGAAA 444

Qy        418 ATAAATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAAT 477
Db        445 ATTGACTATGAAGGATTACTTAGAAATACTATAATAGGATGTTCTACAGTGGTCATTGAT 504

Qy        478 GCCAACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACG---AGGATTATTTGATG 534
Db        505 AAGGAGATTGTAGGAGAGTTCAGCATGCCTTTGGTAAGAAGAGGACAAGATACTGCAACC 564

Qy        535 TGGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTAT 594
Db        565 TGGTTACAGTTACTTAAAAAGGAAAAATATGCATATGGGATTCAAGAAGATTTAGTGAAT 624

Qy        595 TACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGG 654
Db        625 TACAGATTAGTTGGAAATTCTATATCAAGTAACAAGATAAAAGCATTAAAACGAACTTGG 684

Qy        655 AGTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTA 714
Db        685 AATACTTATAGAAATGTTGAAAATTTAAGTTTACCAAAAAGTCTATATGTATTTTGTTTC 744

Qy        715 TATGCTTCAAATGGAGTCATGAAAAAAATA 744
Db        745 TATGTTTTTAATGCCATTAAAAAAAGAGTA 774
```

RESULT 5
US-09-070-927A-125/c

; Sequence 125, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 125:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4915 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 125:
US-09-070-927A-125

Query Match 17.9%; Score 140.4; DB 3; Length 4915;
Best Local Similarity 50.8%; Pred. No. 1.8e-18;
Matches 361; Conservative 0; Mismatches 346; Indels 3; Gaps 1;

```
Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
Db       4211 ATGGAAACAGCACTTGTTTCAATTATTATGCCCATGTACAATGCTGGCAAATTTTTATCA 4152

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
Db       4151 AAATCTATTGAAAGCGTTCTGGAACAAACCTATCAGAATTGGGAATTATTACTCATTGAT 4092

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
Db       4091 GATGGAAGTAAAGATGATTCTATTGATATTGCRCTTGCCTTTATGGAAAAAGATTCAAGA 4032

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
Db       4031 ATCTTTTTACTGAAAAATGAACAGAACATGGGCATTGCAAAGACCAGAAATAAAGGAATT 3972

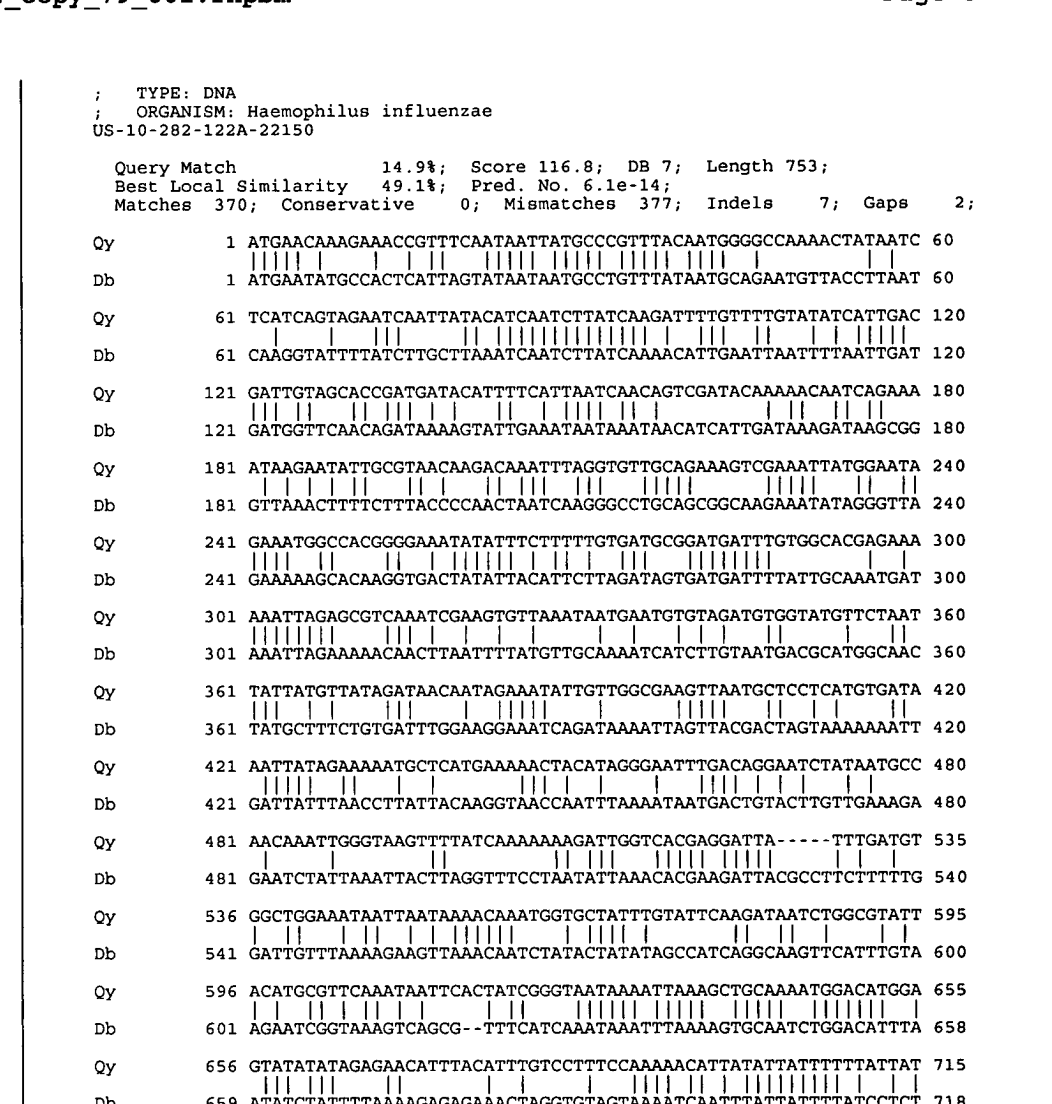
Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
Db       3971 GAAGCTTCAAAGGGACAATATATTGCCTTTTTAGATAGTGATGATTTATGGCTTCCTAAT 3912

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
```



RESULT 6
US-10-282-122A-22150
; Sequence 22150, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22150
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-10-282-122A-22150

Query Match 14.9%; Score 116.8; DB 7; Length 753;
Best Local Similarity 49.1%; Pred. No. 6.1e-14;
Matches 370; Conservative 0; Mismatches 377; Indels 7; Gaps 2;



RESULT 7
US-10-329-670-1
; Sequence 1, Application US/10329670
; Publication No. US20040018503A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragm
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,670
; CURRENT FILING DATE: 2002-12-24
; PRIOR APPLICATION NUMBER: US 09/643,990

; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)

```
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (131360)..(131360)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (139910)..(139910)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (140398)..(140398)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (142750)..(142750)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (145058)..(145058)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (145171)..(145171)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (145942)..(145942)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (147197)..(147197)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (150841)..(150841)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152500)..(152500)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152530)..(152530)

  Query Match             14.9%;  Score 116.8;  DB 7;  Length 1830121;
  Best Local Similarity   49.1%;  Pred. No. 1.1e-12;
  Matches  370;  Conservative    0;  Mismatches  377;  Indels    7;  Gaps    2;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              |||||  |     |   | ||   ||||| ||||| ||||| ||||  |          | |
Db     919974 ATGAATATGCCACTCATTAGTATAATAATGCCTGTTTATAATGCAGAATGTTACCTTAAT 920033

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
                |     |   |||      || ||||||||||||||  |  |||  ||   | | |||||
Db     920034 CAAGGTATTTTATCTTGCTTAAATCAATCTTATCAAAACATTGAATTAATTTTAATTGAT 920093

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||| ||   || ||| | |   ||  | |||| || |      |  ||  || ||
Db     920094 GATGGTTCAACAGATAAAAGTATTGAAATAATAAATAACATCATTGATAAAGATAAGCGG 920153

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              | | | ||   || |   || |||  |||   |||||      |||||   ||  ||
Db     920154 GTTAAACTTTTCTTTACCCCAACTAATCAAGGGCCTGCAGCGGCAAGAAATATAGGGTTA 920213

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||  ||    ||  | |||||| | || |  |||   ||||||||        | |
Db     920214 GAAAAAGCACAAGGTGACTATATTACATTCTTAGATAGTGATGATTTTATTGCAAATGAT 920273

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||    ||| |  |  |  |      |  |   | | |   ||      |   ||
Db     920274 AAATTAGAAAAACAACTTAATTTTATGTTGCAAAATCATCTTGTAATGACGCATGGCAAC 920333

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||  | |    |||     |  |||||     |       |||||   ||  | |    ||
Db     920334 TATGCTTTCTGTGATTTGGAAGGAAATCAGATAAAATTAGTTACGACTAGTAAAAAAATT 920393

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              |||||  ||   |  |        ||| |  |   |   |||| | | |   | |
Db     920394 GATTATTTAACCTTATTACAAGGTAACCAATTTAAAATAATGACTGTACTTGTTGAAAGA 920453

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTA-----TTTGATGT 535
              |      |        ||         || |||   ||||| |||||     | |  |
Db     920454 GAATCTATTAAATTACTTAGGTTTCCTAATATTAAACACGAAGATTACGCCTTCTTTTTG 920513

Qy        536 GGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATT 595
              |  ||    | ||  | | ||||||     | |||| |         ||  ||  |    | |
Db     920514 GATTGTTTAAAAGAAGTTAAACAATCTATACTATATAGCCATCAGGCAAGTTCATTTGTA 920573

Qy        596 ACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGA 655
              | |  || | || | |      | ||    |||||| |||||  |||||  |||||||  |
Db     920574 AGAATCGGTAAAGTCAGCG--TTTCATCAAATAAATTTAAAAGTGCAATCTGGACATTTA 920631

Qy        656 GTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTAT 715
              |||  |||    ||       |  |      |   |||| || | |||||||||  |  | |
Db     920632 ATATCTATTTTAAAAGAGAGAAACTAGGTGTAGTAAAATCAATTTATTATTTTATCCTCT 920691

Qy        716 ATGCTTCAAATGGAGTCATGAAAAAAATAACACA 749
              |||| |  ||||||  | || |||  | |  | | |
Db     920692 ATGCCTATAATGGATTTATTAAATACAAGAAATA 920725


RESULT 8
US-10-158-865-1
; Sequence 1, Application US/10158865
; Publication No. US20040203093A1
; GENERAL INFORMATION:
;  APPLICANT: Fleischmann et al.
;  TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Fragm
;  TITLE OF INVENTION:  Thereof, and Uses Thereof
;  FILE REFERENCE: PB186P2C1D1
;  CURRENT APPLICATION NUMBER: US/10/158,865
;  CURRENT FILING DATE:  2002-06-03
;  PRIOR APPLICATION NUMBER: US 09/557,884
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: US 08/476,102
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: US 08/426,787
;  PRIOR FILING DATE: 1995-04-21
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1830121
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4747)..(4747)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9921)..(9921)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (10150)..(10150)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (29298)..(29298)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36543)..(36543)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
```

; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (147197)..(147197)

```
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (150841)..(150841)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152500)..(152500)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152530)..(152530)

  Query Match            14.9%;  Score 116.8;  DB 8;  Length 1830121;
  Best Local Similarity   49.1%;  Pred. No. 1.1e-12;
  Matches  370;  Conservative    0;  Mismatches  377;  Indels    7;  Gaps    2;

Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||| |    |  | ||   ||||| ||||| ||||| ||||  |          | |
Db     919974 ATGAATATGCCACTCATTAGTATAATAATGCCTGTTTATAATGCAGAATGTTACCTTAAT 920033

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
                |    |  |||     || ||||||||||||| |  |||  ||   | | |||||
Db     920034 CAAGGTATTTTATCTTGCTTAAATCAATCTTATCAAAACATTGAATTAATTTTAATTGAT 920093

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||| ||   || ||| | |   ||  | |||| || |          | ||  || ||
Db     920094 GATGGTTCAACAGATAAAAGTATTGAAATAATAAATAACATCATTGATAAAGATAAGCGG 920153

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              | | | | ||   || |   || |||  |||   |||||       |||||   ||  ||
Db     920154 GTTAAACTTTTCTTTACCCCAACTAATCAAGGGCCTGCAGCGGCAAGAAATATAGGGTTA 920213

Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||  ||    ||  | |||||| | || |  |||   |||||||        |  |
Db     920214 GAAAAAGCACAAGGTGACTATATTACATTCTTAGATAGTGATGATTTTATTGCAAATGAT 920273

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||    ||| |  | |  |      |  |   | | |   ||     |   ||
Db     920274 AAATTAGAAAAACAACTTAATTTTATGTTGCAAAATCATCTTGTAATGACGCATGGCAAC 920333

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||  | |    |||    | |||||    |       |||||   ||  | |    ||
Db     920334 TATGCTTTCTGTGATTTGGAAGGAAATCAGATAAAATTAGTTACGACTAGTAAAAAAATT 920393

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              |||||  ||   | |        ||| | |    |   |||| | | |   | |
Db     920394 GATTATTTAACCTTATTACAAGGTAACCAATTTAAAATAATGACTGTACTTGTTGAAAGA 920453

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTA-----TTTGATGT 535
              |     |       ||          || |||   ||||| |||||     | |  |
Db     920454 GAATCTATTAAATTACTTAGGTTTCCTAATATTAAACACGAAGATTACGCCTTCTTTTTG 920513

Qy        536 GGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATT 595
              |  ||   | ||  | | ||||||    | |||| |       ||  ||  |    | |
Db     920514 GATTGTTTAAAAGAAGTTAAACAATCTATACTATATAGCCATCAGGCAAGTTCATTTGTA 920573

Qy        596 ACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGA 655
              | |  || | || | |     | ||    |||||| ||||  |||||  |||||||  |
Db     920574 AGAATCGGTAAAGTCAGCG--TTTCATCAAATAAATTTAAAAGTGCAATCTGGACATTTA 920631

Qy        656 GTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTAT 715
              |||  |||    ||       |  |     |   |||| || | |||||||||  |  | |
Db     920632 ATATCTATTTTAAAAGAGAGAAACTAGGTGTAGTAAAATCAATTTATTATTTTATCCTCT 920691

Qy        716 ATGCTTCAAATGGAGTCATGAAAAAAATAACACA 749
              |||| |  |||||| | || ||| | |   | | |
Db     920692 ATGCCTATAATGGATTTATTAAATACAAGAAATA 920725


RESULT 9
US-10-981-687-1
; Sequence 1, Application US/10981687

; Publication No. US20050131222A1
; GENERAL INFORMATION:
;  APPLICANT: Fleischmann et al.
;  TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Fragm
;  TITLE OF INVENTION:  Thereof, and Uses Thereof
;  FILE REFERENCE: PB186P2C1D12
;  CURRENT APPLICATION NUMBER: US/10/981,687
;  CURRENT FILING DATE:  2004-11-05
;  PRIOR APPLICATION NUMBER: US 10/158,865
;  PRIOR FILING DATE: 2002-06-03
;  PRIOR APPLICATION NUMBER: US 09/557,884
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: US 08/476,102
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: US 08/426,787
;  PRIOR FILING DATE: 1995-04-21
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1830121
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4747)..(4747)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9921)..(9921)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (10150)..(10150)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (29298)..(29298)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36543)..(36543)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36551)..(36551)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36636)..(36636)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (40808)..(40810)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (44416)..(44416)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (44905)..(44905)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (44975)..(44975)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (45593)..(45593)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (45732)..(45732)
```

; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:

Query Match 14.9%; Score 116.8; DB 9; Length 1830121;
Best Local Similarity 49.1%; Pred. No. 1.1e-12;
Matches 370; Conservative 0; Mismatches 377; Indels 7; Gaps 2;

```
Qy          1 ATGAACAAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATC 60
              ||||| |     |  | ||   ||||| ||||| ||||| ||||  |        | ||
Db     919974 ATGAATATGCCACTCATTAGTATAATAATGCCTGTTTATAATGCAGAATGTTACCTTAAT 920033

Qy         61 TCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGAC 120
                |     |   |||      || ||||||||||||| |  |||  ||   | | |||||
Db     920034 CAAGGTATTTTATCTTGCTTAAATCAATCTTATCAAAACATTGAATTAATTTTAATTGAT 920093

Qy        121 GATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAA 180
              ||| ||   || ||| | |   ||  | |||| || |          | ||  || ||
Db     920094 GATGGTTCAACAGATAAAAGTATTGAAATAATAAATAACATCATTGATAAAGATAAGCGG 920153

Qy        181 ATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATA 240
              | | | | ||   || |   || |||  |||   |||||       |||||   ||  ||
Db     920154 GTTAAACTTTTCTTTACCCCAACTAATCAAGGGCCTGCAGCGGCAAGAAATATAGGGTTA 920213
```

```
Qy        241 GAAATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAA 300
              ||||  ||    ||  | |||||| | || |  |||   ||||||||        |  |
Db     920214 GAAAAAGCACAAGGTGACTATATTACATTCTTAGATAGTGATGATTTTATTGCAAATGAT 920273

Qy        301 AAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAAT 360
              ||||||||    ||| |  |  |  |      |  |    | | |   ||      |   ||
Db     920274 AAATTAGAAAAACAACTTAATTTTATGTTGCAAAATCATCTTGTAATGACGCATGGCAAC 920333

Qy        361 TATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATA 420
              |||  | |    |||    |  |||||     |       |||||   ||  | |    ||
Db     920334 TATGCTTTCTGTGATTTGGAAGGAAATCAGATAAAATTAGTTACGACTAGTAAAAAAATT 920393

Qy        421 AATTATAGAAAAATGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCC 480
              |||||  ||   |  |       ||| |  |    |   |||| | | |   | |
Db     920394 GATTATTTAACCTTATTACAAGGTAACCAATTTAAAATAATGACTGTACTTGTTGAAAGA 920453

Qy        481 AACAAATTGGGTAAGTTTTATCAAAAAAAGATTGGTCACGAGGATTA-----TTTGATGT 535
              |     |         ||          || |||   ||||| |||||      | | |
Db     920454 GAATCTATTAAATTACTTAGGTTTCCTAATATTAAACACGAAGATTACGCCTTCTTTTTG 920513

Qy        536 GGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATT 595
              |  ||   | ||  | | ||||||    | |||| |       ||  ||  |    | |
Db     920514 GATTGTTTAAAAGAAGTTAAACAATCTATACTATATAGCCATCAGGCAAGTTCATTTGTA 920573

Qy        596 ACATGCGTTCAAATAATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGA 655
              | |  || | || | |     | ||    ||||| |||||  |||||  |||||||  |
Db     920574 AGAATCGGTAAAGTCAGCG--TTTCATCAAATAAATTTAAAAGTGCAATCTGGACATTTA 920631

Qy        656 GTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTAT 715
              |||  |||     ||        |  |       |   |||| || | |||||||||  |  | |
Db     920632 ATATCTATTTTAAAAGAGAGAAACTAGGTGTAGTAAAATCAATTTATTATTTTATCCTCT 920691

Qy        716 ATGCTTCAAATGGAGTCATGAAAAAAATAACACA 749
              |||| |  |||||| | || ||| | |  | | |
Db     920692 ATGCCTATAATGGATTTATTAAATACAAGAAATA 920725


RESULT 10
US-09-816-028A-30
; Sequence 30, Application US/09816028A
; Patent No. US20020042369A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/09/816,028A
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
;   LENGTH: 891
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(891)
;   OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
US-09-816-028A-30

  Query Match            12.8%;  Score 100;  DB 3;  Length 891;
  Best Local Similarity   48.5%;  Pred. No. 1.7e-10;
  Matches  333;  Conservative    0;  Mismatches  349;  Indels    4;  Gaps    2;

Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76
              |||||||  ||  ||||   ||| |||| ||   ||  ||||       || |||||

Db         11 TTTCAATCATCTTGCCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGTT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136
              |||  ||||  ||| | || || | |    |   | |  |||| ||||||  | |  ||
Db         71 GTATCAATCAGACTTTTAAAAATATAGAAATAATTGTAGTTGATGATTGTGGAAGTGACA 130

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
              | |   |     || | ||          ||||||  || | | ||||| |||  | | ||
Db        131 AAAGTATAGATATAGTTAAAGAATATGCCAAAAAAGATGATAGAATAAAAATCATACATA 190

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256
              |  |  |||||||    ||  |  |  | || || | ||  || || | || |
Db        191 ATGAAGAAAATTTAAAACTTTTAAGAGCTAGATATGAAGGTGTAAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316
              |||||    ||||  ||| | |||||||  |   |    ||     |  ||       |
Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTAGAACTTAATGCTTGTGAAGAATGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376
              |  || | |||||  ||  |     ||||| |  ||| || ||||   | ||| | || |
Db        311 TGAAAATTTTAAAAAACAATGAAATAGATTTATTATTTTTTAATGCATTTGTATTGGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436
              | || |  || || |    |  |  | ||   |  ||| |  | |  ||||  |||||
Db        371 ATAACAATAAAATAGAAAGAAAGTTGAA---TTTTCAAGAAAAATGTTATGTAAAAAAAG 427

Qy        437 TCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTGGGTAAGT 496
              |   |||  |  ||   ||    | |    | | |    |
Db        428 ATTTTTTAAAAGAACTATTAAAAACTAAAAATTTATTTTGGACAGTGTGGGCAAAAGTCA 487

Qy        497 TTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGGAAATAATTAATAAAA 556
              | |  || ||      |  | |  | |  ||| || | ||| ||||    ||| | ||
Db        488 TAAAAAAAGAATTATATCTCAAGGCTGTTGGTTTAATATCGCTAGAAAATGCTAAAATAA 547

Qy        557 CAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCAAATAATTCAC 616
              |  | ||     |  | |  | || | |     |    | ||        | ||  |  |
Db        548 ATATGGCTGAAGATGTTTTATTATATTACCCTTTGATAAATATTTCAAATACTATATTTC 607

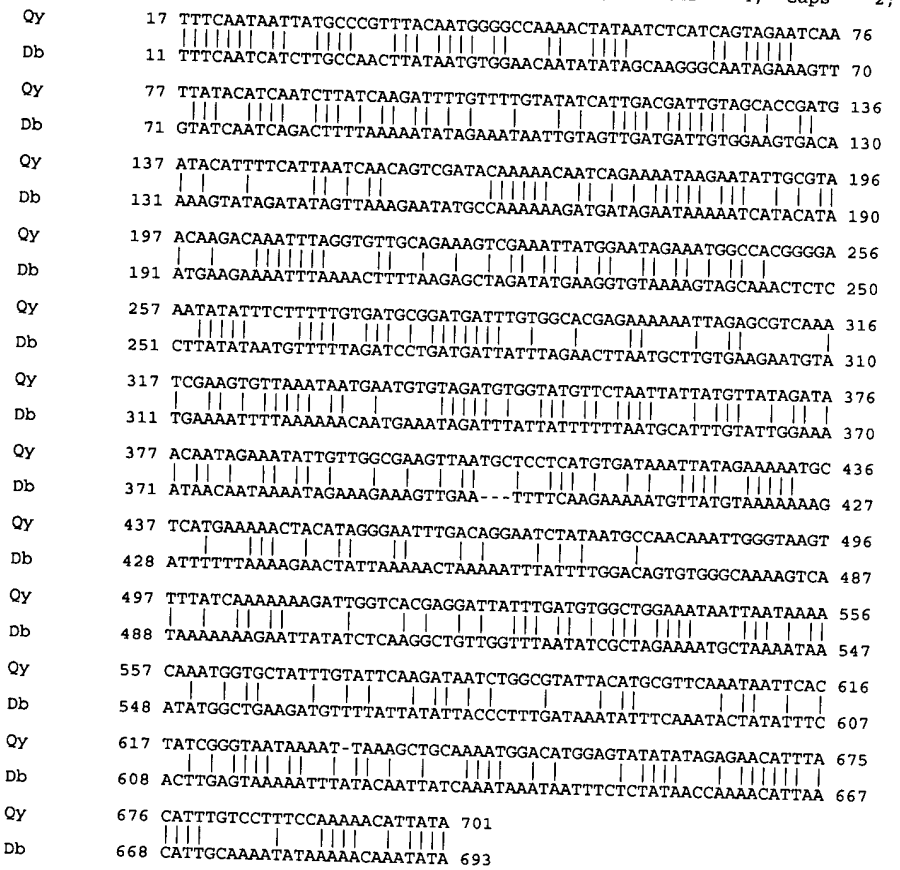
Qy        617 TATCGGGTAATAAAAT-TAAAGCTGCAAAATGGACATGGAGTATATATAGAGAACATTTA 675
              | | |||| ||  | || |  |   ||||  | |     | ||||   | ||||||
Db        608 ACTTGAGTAAAAATTTATACAATTATCAAATAAATAATTTCTCTATAACCAAAACATTAA 667

Qy        676 CATTTGTCCTTTCCAAAAACATTATA 701
              ||||      |   ||||  | ||||
Db        668 CATTGCAAAATATAAAAACAAATATA 693


RESULT 11
US-10-303-161-30
; Sequence 30, Application US/10303161
; Publication No. US20030148459A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,161
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
;   LENGTH: 891
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
```

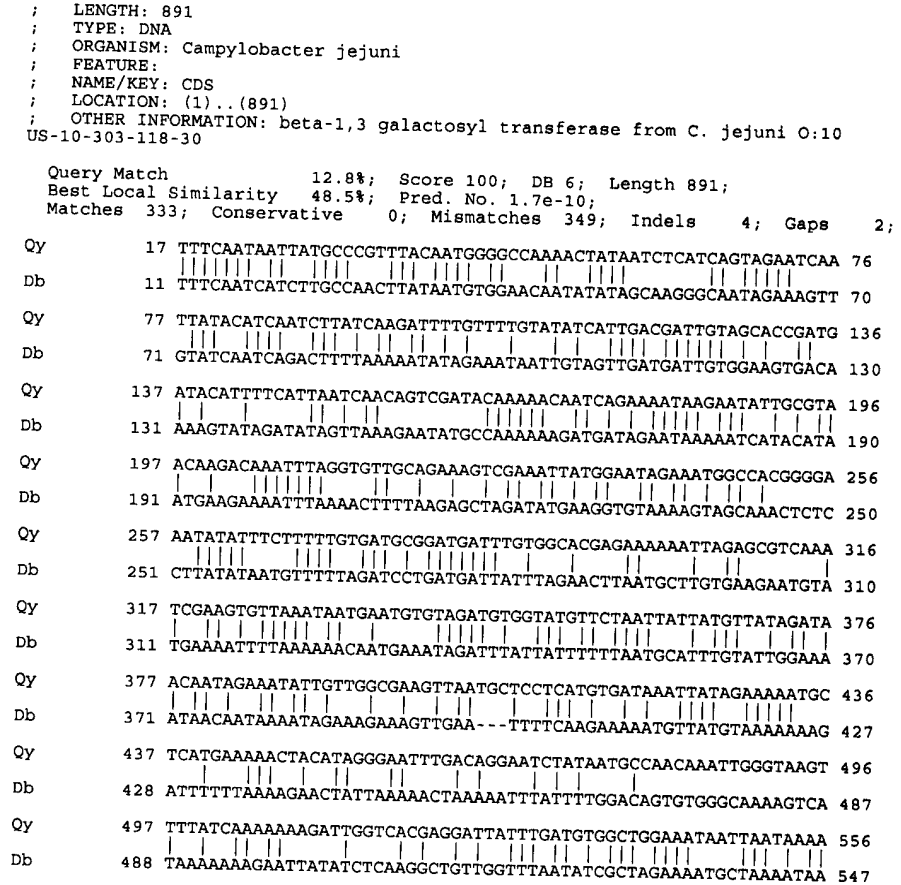
; NAME/KEY: CDS
; LOCATION: (1)..(891)
; OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
US-10-303-161-30

Query Match 12.8%; Score 100; DB 6; Length 891;
Best Local Similarity 48.5%; Pred. No. 1.7e-10;
Matches 333; Conservative 0; Mismatches 349; Indels 4; Gaps 2;



RESULT 12
US-10-303-118-30
; Sequence 30, Application US/10303118
; Publication No. US20030157655A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,118
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(891)
; OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
US-10-303-118-30

Query Match 12.8%; Score 100; DB 6; Length 891;
Best Local Similarity 48.5%; Pred. No. 1.7e-10;
Matches 333; Conservative 0; Mismatches 349; Indels 4; Gaps 2;



RESULT 13
US-10-303-128-30
; Sequence 30, Application US/10303128
; Publication No. US20030157656A1

```
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,128
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
;   LENGTH: 891
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(891)
;   OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
US-10-303-128-30

  Query Match             12.8%;  Score 100;  DB 6;  Length 891;
  Best Local Similarity   48.5%;  Pred. No. 1.7e-10;
  Matches  333;  Conservative    0;  Mismatches  349;  Indels    4;  Gaps    2;

Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76
              ||||||| ||  ||||    ||| |||| ||    ||  ||||        || |||||
Db         11 TTTCAATCATCTTGCCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGTT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136
                |||  ||||  ||| | || || | |   |  | |  |||| |||||| | |  ||
Db         71 GTATCAATCAGACTTTTAAAAATATAGAAATAATTGTAGTTGATGATTGTGGAAGTGACA 130

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
              | |  |     || | ||          ||||||  || | | ||||| |||  | | ||
Db        131 AAAGTATAGATATAGTTAAAGAATATGCCAAAAAAGATGATAGAATAAAAATCATACATA 190

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256
              |  |  |||||||    ||  |  |  | || || | ||  || || | | || |
Db        191 ATGAAGAAAATTTAAAACTTTTAAGAGCTAGATATGAAGGTGTAAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316
                |||||    ||||  ||| | |||||||  |   |    ||    |  ||       |
Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTAGAACTTAATGCTTGTGAAGAATGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376
              |  || | ||||| ||  |    |||||  |  ||| || ||||   | |||  | || |
Db        311 TGAAAATTTTAAAAAACAATGAAATAGATTTATTATTTTTTAATGCATTTGTATTGGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436
              | || |  || || |   |  | ||   |  ||| |  | |  ||||  |||||
Db        371 ATAACAATAAAATAGAAAGAAAGTTGAA---TTTTCAAGAAAAATGTTATGTAAAAAAAG 427

Qy        437 TCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTGGGTAAGT 496
                |   |||  |  ||   ||    | |    | | |    |
Db        428 ATTTTTTAAAAGAACTATTAAAAACTAAAAATTTATTTTGGACAGTGTGGGCAAAAGTCA 487

Qy        497 TTTATCAAAAAAAGATTGGTCACGAGGATTATTTGATGTGGCTGGAAATAATTAATAAAA 556
              |  |  || ||     |   | |  | |  ||| || | ||| ||||    ||| | ||
Db        488 TAAAAAAAGAATTATATCTCAAGGCTGTTGGTTTAATATCGCTAGAAAATGCTAAAATAA 547

Qy        557 CAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCAAATAATTCAC 616
              |  | ||    |  | |   | || | |    |    | ||        | ||  |  |
Db        548 ATATGGCTGAAGATGTTTTATTATATTACCCTTTGATAAATATTTCAAATACTATATTTC 607

Qy        617 TATCGGGTAATAAAAT-TAAAGCTGCAAAATGGACATGGAGTATATATAGAGAACATTTA 675
                | | |||| ||  | || | |  |||| | |    | ||||   | |||||| |
```

```
Db        608 ACTTGAGTAAAAATTTATACAATTATCAAATAAATAATTTCTCTATAACCAAAACATTAA 667

Qy        676 CATTTGTCCTTTCCAAAAACATTATA 701
                ||||      |  ||||  | ||||
Db        668 CATTGCAAAATATAAAAACAAATATA 693


RESULT 14
US-10-303-134-30
; Sequence 30, Application US/10303134
; Publication No. US20030157657A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,134
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
;   LENGTH: 891
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(891)
;   OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
US-10-303-134-30

  Query Match             12.8%;  Score 100;  DB 6;  Length 891;
  Best Local Similarity   48.5%;  Pred. No. 1.7e-10;
  Matches  333;  Conservative    0;  Mismatches  349;  Indels    4;  Gaps    2;

Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76
              ||||||| ||  ||||    ||| |||| ||    ||  ||||        || |||||
Db         11 TTTCAATCATCTTGCCAACTTATAATGTGGAACAATATATAGCAAGGGCAATAGAAAGTT 70

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136
                |||  ||||  ||| | || || | |   |  | |  |||| |||||| | |  ||
Db         71 GTATCAATCAGACTTTTAAAAATATAGAAATAATTGTAGTTGATGATTGTGGAAGTGACA 130

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
              | |  |     || | ||          ||||||  || | | ||||| |||  | | ||
Db        131 AAAGTATAGATATAGTTAAAGAATATGCCAAAAAAGATGATAGAATAAAAATCATACATA 190

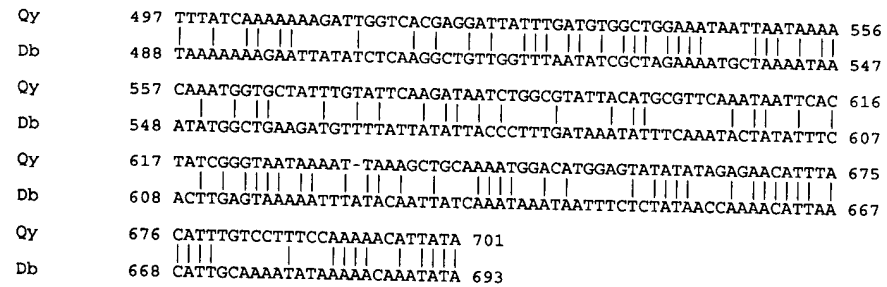
Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256
              |  |  |||||||    ||  |  |  | || || | ||  || || | | || |
Db        191 ATGAAGAAAATTTAAAACTTTTAAGAGCTAGATATGAAGGTGTAAAAGTAGCAAACTCTC 250

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316
                |||||    ||||  ||| | |||||||  |   |    ||    |  ||       |
Db        251 CTTATATAATGTTTTTAGATCCTGATGATTATTTAGAACTTAATGCTTGTGAAGAATGTA 310

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376
              |  || | ||||| ||  |    |||||  |  ||| || ||||   | |||  | || |
Db        311 TGAAAATTTTAAAAAACAATGAAATAGATTTATTATTTTTTAATGCATTTGTATTGGAAA 370

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAAAAATGC 436
              | || |  || || |   |  | ||   |  ||| |  | |  ||||  |||||
Db        371 ATAACAATAAAATAGAAAGAAAGTTGAA---TTTTCAAGAAAAATGTTATGTAAAAAAAG 427

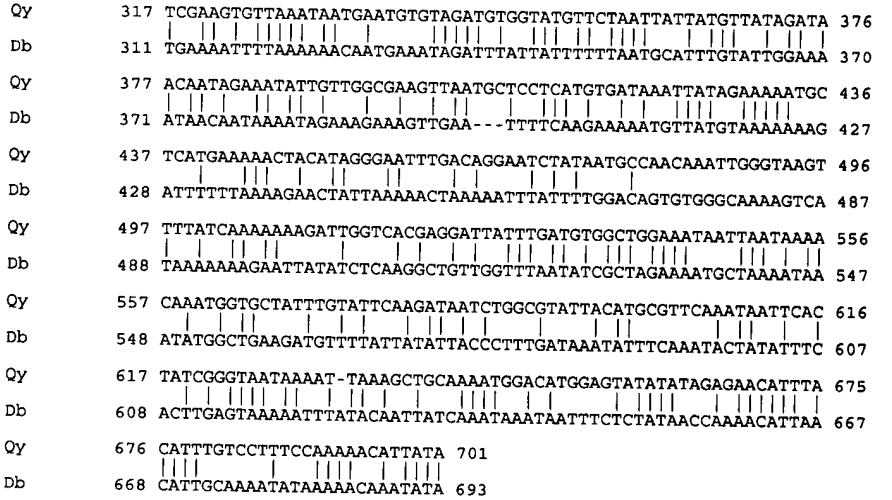
Qy        437 TCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTGGGTAAGT 496
                |   |||  |  ||   ||    | |    | | |    |
Db        428 ATTTTTTAAAAGAACTATTAAAAACTAAAAATTTATTTTGGACAGTGTGGGCAAAAGTCA 487
```



RESULT 15
US-10-303-162-30
; Sequence 30, Application US/10303162
; Publication No. US20030157658A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,162
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(891)
; OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
US-10-303-162-30

Query Match 12.8%; Score 100; DB 6; Length 891;
Best Local Similarity 48.5%; Pred. No. 1.7e-10;
Matches 333; Conservative 0; Mismatches 349; Indels 4; Gaps 2;





Search completed: May 11, 2006, 07:08:17
Job time : 524.803 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:23:58 ; Search time 404.616 Seconds
(without alignments)
7892.621 Million cell updates/sec

Title: US-09-423-093B-2_COPY_79_861
Perfect score: 783
Sequence: 1 atgaacaaagaaaccgtttc.....gaaaggagactaaaaagtga 783

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	70	8.9	891	18	US-11-079-463-838	Sequence 838, App
	2	69.4	8.9	930	18	US-11-079-463-1998	Sequence 1998, Ap
	3	69.2	8.8	1774	10	US-10-510-386-103	Sequence 103, App
	4	67.2	8.6	765	18	US-11-079-463-2016	Sequence 2016, Ap
	5	67	8.6	1851	17	US-11-172-145-5	Sequence 5, Appli
	6	67	8.6	1854	17	US-11-172-145-7	Sequence 7, Appli
	7	66	8.4	1146	17	US-11-098-686-9829	Sequence 9829, Ap
c	8	66	8.4	1457619	17	US-11-098-686-8739	Sequence 8739, Ap
	9	59	7.5	2944528	18	US-11-045-004-1	Sequence 1, Appli
	10	57.4	7.3	999	18	US-11-079-463-1654	Sequence 1654, Ap
c	11	56	7.2	2736	17	US-11-194-246-190	Sequence 190, App
	12	55	7.0	2979	17	US-11-172-145-3	Sequence 3, Appli
	13	54.2	6.9	723	18	US-11-079-463-576	Sequence 576, App
	14	53.4	6.8	897	18	US-11-079-463-1041	Sequence 1041, Ap
c	15	51.8	6.6	2944528	18	US-11-045-004-1	Sequence 1, Appli
	16	51.2	6.5	1020	17	US-11-098-686-8753	Sequence 8753, Ap
c	17	51.2	6.5	27048	17	US-11-098-686-8736	Sequence 8736, Ap
	18	50.6	6.5	998	12	US-10-301-480-613360	Sequence 613360,
	19	50.6	6.5	998	12	US-10-301-480-1226769	Sequence 1226769,
	20	50.4	6.4	26772	10	US-10-995-561-13313	Sequence 13313, A
	21	50.4	6.4	54946	10	US-10-995-561-13479	Sequence 13479, A
	22	50.2	6.4	1764	17	US-11-172-145-16	Sequence 16, Appl
	23	50.2	6.4	1830	17	US-11-172-145-15	Sequence 15, Appl
	24	50.2	6.4	1902	17	US-11-172-145-14	Sequence 14, Appl
	25	50.2	6.4	1980	17	US-11-172-145-13	Sequence 13, Appl
	26	50.2	6.4	2112	17	US-11-172-145-11	Sequence 11, Appl
	27	50.2	6.4	2112	17	US-11-172-145-17	Sequence 17, Appl
	28	50.2	6.4	2112	17	US-11-172-145-18	Sequence 18, Appl
	29	50.2	6.4	2920	17	US-11-172-145-1	Sequence 1, Appli
	30	50.2	6.4	2937	17	US-11-124-215-2	Sequence 2, Appli
	31	50.2	6.4	2937	17	US-11-120-422-9	Sequence 9, Appli
	32	50.2	6.4	2937	17	US-11-226-480-8	Sequence 8, Appli
	33	50.2	6.4	2937	17	US-11-228-079-8	Sequence 8, Appli
	34	49.8	6.4	2990	10	US-10-793-626-3899	Sequence 3899, Ap
c	35	49.4	6.3	687411	11	US-10-330-773-26	Sequence 26, Appl
	36	49.2	6.3	1152	18	US-11-079-463-1593	Sequence 1593, Ap
	37	49	6.3	951	18	US-11-079-463-497	Sequence 497, App
	38	48.6	6.2	2112	17	US-11-172-145-19	Sequence 19, Appl
	39	48.6	6.2	2112	17	US-11-172-145-20	Sequence 20, Appl
	40	48.6	6.2	49979	10	US-10-995-561-13443	Sequence 13443, A
	41	47.6	6.1	554	12	US-10-301-480-378566	Sequence 378566,
	42	47.6	6.1	554	12	US-10-301-480-991975	Sequence 991975,
	43	47.6	6.1	562	7	US-09-925-065A-302348	Sequence 302348,
c	44	47.6	6.1	49979	10	US-10-995-561-13443	Sequence 13443, A
	45	47.4	6.1	6583	10	US-10-240-708-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-11-079-463-838
; Sequence 838, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAC
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 838
; LENGTH: 891
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-838

Query Match 8.9%; Score 70; DB 18; Length 891;
Best Local Similarity 50.6%; Pred. No. 0.0024;
Matches 169; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

```
Qy         13 ACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAA 72
              ||  |  || |  ||| |||     |  ||| ||||||       ||     || |   |||
Db         10 ACTATAACAGTGTTTACGCCGACATTTAATCGGGCCACTCTGTTACCTAGATTATATGAA 69

Qy         73 TCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACC 132
                 |  |||||  | | ||  |||||||  | |  |||  | || ||| | || ||
Db         70 AGTTTAGTTAATCAAACATTTCTTGATTTTGAGTGGCTTATTGTAGATGATGGAAGTACT 129

Qy        133 GATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTG 192
              ||||||||||||   || ||  |         |  ||||| || |||     | |  ||
Db        130 GATGATACATTTAACTTTATAGAATCAATTAAGGAAAACGATAAGATTGATATACAATAT 189
```

```
Qy        193 CGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACG 252
Db        190 TACTATCAGAATAATGCTGGTAAACATGCTGCTATTAATTGGGGAGTAGAGCTAGCTAAA 249

Qy        253 GGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGT 312
Db        250 GGCGATCTTTTTTTTATTGTAGATAGTGATGAGGTTATGATTGAATCTGGGTTACAGACT 309

Qy        313 CAAATCGAAGTGTTAAATAATGAATGTGTAGATG 346
Db        310 ATAGTAGATGTTTATAAACAAGTATCTGATAATG 343
```

RESULT 2
US-11-079-463-1998
; Sequence 1998, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 1998
; LENGTH: 930
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-1998

Query Match 8.9%; Score 69.4; DB 18; Length 930;
Best Local Similarity 49.3%; Pred. No. 0.0031;
Matches 181; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

```
Qy         16 GTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCA 75
Db         31 GTATCTATAGTAATTCCGGTATATAACAGAGAACTATACATAGAAGATGCTATAAAATCT 90

Qy         76 ATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGAT 135
Db         91 GCTATTAGTCAAACGTATCAAAATATTGAGATTATTATAGTTGATAACTGTAGTACAGAT 150

Qy        136 GATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGT 195
Db        151 TTAACTTGGAGTATTTTGAATACATGGGCAAGCAAAGATCATAGAATTAAAATATATCAG 210

Qy        196 AACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGG 255
Db        211 AATAAAACTAATATAGGACCTGTTTTAAATTGGAATGAATGCTTTAAGCATGCTTCAGGT 270

Qy        256 AAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAA 315
Db        271 GAGTATATAAAGATTCTTTGGTCTGATGATTGGATGGCATTAGATTTTATTGAAAGAGCA 330

Qy        316 ATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGAT 375
Db        331 GTGAATCTAATTGATGAGAAGTCAGCGTTTGTTATATCAAATCATAAGATCGTATCAGAA 390

Qy        376 AACAATA 382
Db        391 AATGGTA 397
```

RESULT 3
US-10-510-386-103
; Sequence 103, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 103
; LENGTH: 1774
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (516)..(1271)
US-10-510-386-103

Query Match 8.8%; Score 69.2; DB 10; Length 1774;
Best Local Similarity 45.3%; Pred. No. 0.0036;
Matches 334; Conservative 0; Mismatches 398; Indels 6; Gaps 2;

```
Qy         17 TTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAA 76
Db        538 TCTCTGTCATTACGCCTTCCTATAACGCGGAGGAATTTATCGAAAAAACCATTAAGTCCG 597

Qy         77 TTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATG 136
Db        598 TCTTAAATCAGACGTTTTCCGACTGGGAAATGATCATTGCCGATGATTGCTCAACAGATG 657

Qy        137 ATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTA 196
Db        658 GGACAAGAGACATTTTGAAGCGCTATGAAGAAGATGATGAGCGGATCCATGCCATTTTTC 717

Qy        197 ACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGA 256
Db        718 TGAAAGAGAATCAAGGTGCTGCAGCGGCGCGAAATGCGGCGCTCAGCAAAGCCGAAGGGC 777

Qy        257 AATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAA 316
Db        778 GCTATGTCGCCTTTTTGGACAGCGATGATGTTTGGAAAGCAGAAAAGCTGGATAAGCAGC 837

Qy        317 TCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATA 376
Db        838 TCGCCTTCATGAGAAAGCATCAGCATGCTTTCTCATTTACGGCATACGAGCTGATCAGCC 897

Qy        377 ACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGA---TAAATTATAGAAAAA 433
Db        898 AAGACGGCGAACCGCTTCATAAAACCATTCATGCACCCGTAAGCCTTACATATGATGATG 957

Qy        434 TGCTCATGAAAAACTACATAGGGAATTTGACAGGAATCTATAATGCCAACAAATTGGGTA 493
Db        958 TATTAAAAAATACGATCATCGGCTGCTTGACGGTGATGATTGACAGAGAACAAACGGGTG 1017

Qy        494 AGTTTTATCAAAAAAAGATTGG---TCACGAGGATTATTTGATGTGGCTGGAAATAATTA 550
Db       1018 ATATCCGGATGCCGAATATTAGAACCCGCCAGGATTTGGCGACATGGCTGTCCGTATTAA 1077

Qy        551 ATAAAACAAATGGTGCTATTTGTATTCAAGATAATCTGGCGTATTACATGCGTTCAAATA 610
Db       1078 AGCGGGGGTTCAAGGCATATGGACTGAACGAACCCCTCGCGGAATACCGCATCGTTGAAA 1137

Qy        611 ATTCACTATCGGGTAATAAAATTAAAGCTGCAAAATGGACATGGAGTATATATAGAGAAC 670
Db       1138 CATCGATCTCCAGAAACAAGTGGAAGGCGGCGCGAAAAACCTGGTACGTATACAGGGAAA 1197

Qy        671 ATTTACATTTGTCCTTTCCAAAAACATTATATTATTTTTTATTATATGCTTCAAATGGAG 730
Db       1198 TCGAACGACTTCACTTGATGAAAGCGACATGGTGCTTTTTCCATTACGCTAAGAACGCAG 1257
```

```
Qy        731 TCATGAAAAAAATAACAC 748
              | ||||||| | ||  ||
Db       1258 TAATGAAAAGATTATAAC 1275
```

RESULT 4
US-11-079-463-2016
; Sequence 2016, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 2016
; LENGTH: 765
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-2016

Query Match 8.6%; Score 67.2; DB 18; Length 765;
Best Local Similarity 50.5%; Pred. No. 0.0074;
Matches 190; Conservative 0; Mismatches 183; Indels 3; Gaps 1;

```
Qy         16 GTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCA 75
              ||||| ||||| | | ||   || |||  ||    |||  ||    |  |  |||| ||
Db         28 GTTTCGATAATAACGGCCTCATATAATTCGGAAGCAACATTAGAGGCTACTATAGAGTCC 87

Qy         76 ATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGAT 135
              |    ||||  |||||||||||| |||  |    ||| ||||| | | |    | |||
Db         88 GTATACAATCAGACTTATCAAGATATTGAATATATTATAATTGATGGTGGGTCAAGTGAT 147

Qy        136 GATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGT 195
              | ||| | ||| ||||  | |      |          |    | || |
Db        148 AAAACAGTATCAATAATACAAAAGTTTGAACCTTTGTTTTCAGGGCGATTGAAATGGATA 207

Qy        196 AACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGG 255
              |  |  ||  |||  ||   ||   | | ||| | ||  | |||||  || || ||
Db        208 TCAGAACCCGATAAAGGGATTTACGACGCTTGGAATAAGGGGTTGGAAATTGCAACAGGT 267

Qy        256 AAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAA 315
              |||  ||| |||||  ||      |||||| | |          |    || |     |
Db        268 GAATGGATTGCTTTTCTTGGAAGCGATGATATTTATTTACCCGACTCTATACATAAATAT 327

Qy        316 ATCGAAGTGTTAAAT---AATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATA 372
              ||  | || || | |   ||  |||||||| |  |  |||  |||||  | | |  |  ||
Db        328 ATGAATGTTTTGATTTCAAAAAAATGTGTAAACTTTATATCATCTAAGTGTAAATTAGTA 387

Qy        373 GATAACAATAGAAATA 388
              |||   |   ||| |
Db        388 AATAGTGAATTAAAAA 403

RESULT 5
US-11-172-145-5
; Sequence 5, Application US/11172145
; Publication No. US20050272696A1
; GENERAL INFORMATION:
; APPLICANT: DeAngelis, Paul
; TITLE OF INVENTION: METHODS OF SELECTIVELY TREATING DISEASES WITH SPECIFIC
; TITLE OF INVENTION: GLYCOSAMINOGLYCAN POLYMERS
; FILE REFERENCE: 3554.104
; CURRENT APPLICATION NUMBER: US/11/172,145
; CURRENT FILING DATE: 2005-06-30
; PRIOR APPLICATION NUMBER: 60/584,442
; PRIOR FILING DATE: 2004-06-30
; PRIOR APPLICATION NUMBER: 10/642,248
; PRIOR FILING DATE: 2003-08-15
; PRIOR APPLICATION NUMBER: 60/404,356
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 60/479,432
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: 60/491,362
; PRIOR FILING DATE: 2003-07-31
; PRIOR APPLICATION NUMBER: 10/195,908
; PRIOR FILING DATE: 2002-07-15
; PRIOR APPLICATION NUMBER: 09/437,277
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: 60/107,929
; PRIOR FILING DATE: 1998-11-11
; PRIOR APPLICATION NUMBER: 09/283,402
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/080,414
; PRIOR FILING DATE: 1998-04-02
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5
; LENGTH: 1851
; TYPE: DNA
; ORGANISM: Pasteurella multocida
US-11-172-145-5

Query Match 8.6%; Score 67; DB 17; Length 1851;
Best Local Similarity 49.9%; Pred. No. 0.0088;
Matches 197; Conservative 0; Mismatches 195; Indels 3; Gaps 1;

```
Qy   22 ATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAATTATA 81
              || || ||| | | | ||| | ||| || | ||| | | ||| | ||
Db   289 ATTATAATGACTTCTCATAATACAGAAAAATTCATTGAAGCCTCAATTAATTCACTATTA 348

Qy   82 CATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATGATACA 141
              ||| | || | | || | | | | || |||| |||||| ||| | |||
Db   349 TTGCAAACATACAATAACTTAGAAGTTATCGTTGTAGATGATTATAGCACAGATAAAACA 408

Qy   142 TTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTAACAAG 201
              ||| | ||| | || | ||| ||| || ||| || ||
Db   409 TTTCAGATCGCATCCAGAATAGCAAACTCTACAAGTAAAGTAAAAACATTCCGATTAAAC 468

Qy   202 ACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGAAATAT 261
              ||||| |||| |||| ||||| || | | | || | |
Db   469 TCAAATCTAGGGACATACTTTGCGAAAAATACAGGAATTTTAAAGTCTAAAGGAGATATT 528

Qy   262 ATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGA---GCGTCAAATC 318
              |||| ||| ||| |||||| | || ||| | ||| | | || || |
Db   529 ATTTTCTTTCAGGATAGCGATGATGTATGTCACCATGAAAGAATCGAAAGATGTGTTAAT 588

Qy   319 GAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATAAC 378
              | | | ||| ||| || | | || | | | |||| | ||| ||
Db   589 GCATTATTATCGAATAAAGATAATATAGCTGTTAGATGTGCATATTCTAGAATAAATCTA 648

Qy   379 AATAGAAATATTGTTGGCGAAGTTAATGCTCCTCA 413
              | | | | | | | |||||||| | | |
Db   649 GAAACACAAAATATAATAAAAGTTAATGATAATAA 683

RESULT 6
US-11-172-145-7
; Sequence 7, Application US/11172145
; Publication No. US20050272696A1
; GENERAL INFORMATION:
; APPLICANT: DeAngelis, Paul
; TITLE OF INVENTION: METHODS OF SELECTIVELY TREATING DISEASES WITH SPECIFIC
; TITLE OF INVENTION: GLYCOSAMINOGLYCAN POLYMERS
; FILE REFERENCE: 3554.104
; CURRENT APPLICATION NUMBER: US/11/172,145
; CURRENT FILING DATE: 2005-06-30
; PRIOR APPLICATION NUMBER: 60/584,442
; PRIOR FILING DATE: 2004-06-30

; PRIOR APPLICATION NUMBER: 10/642,248
; PRIOR FILING DATE: 2003-08-15
; PRIOR APPLICATION NUMBER: 60/404,356
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 60/479,432
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: 60/491,362
; PRIOR FILING DATE: 2003-07-31
; PRIOR APPLICATION NUMBER: 10/195,908
; PRIOR FILING DATE: 2002-07-15
; PRIOR APPLICATION NUMBER: 09/437,277
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: 60/107,929
; PRIOR FILING DATE: 1998-11-11
; PRIOR APPLICATION NUMBER: 09/283,402
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/080,414
; PRIOR FILING DATE: 1998-04-02
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 7
; LENGTH: 1854
; TYPE: DNA
; ORGANISM: Pasteurella multocida
US-11-172-145-7

Query Match 8.6%; Score 67; DB 17; Length 1854;
Best Local Similarity 49.9%; Pred. No. 0.0088;
Matches 197; Conservative 0; Mismatches 195; Indels 3; Gaps 1;

```
Qy         22 ATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAATTATA 81
              || || ||| |   | | |||   | |||  ||    | ||| |  | ||| |  ||
Db        289 ATTATAATGACTTCTCATAATACAGAAAAATTCATTGAAGCCTCAATTAATTCACTATTA 348

Qy         82 CATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATGATACA 141
                 ||| | ||  |  | || |   |    | | || |||| |||||| ||| | |||
Db        349 TTGCAAACATACAATAACTTAGAAGTTATCGTTGTAGATGATTATAGCACAGATAAAACA 408

Qy        142 TTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTAACAAG 201
              |||    |    |||   |  ||     |  ||| ||| || ||| ||    ||
Db        409 TTTCAGATCGCATCCAGAATAGCAAACTCTACAAGTAAAGTAAAAACATTCCGATTAAAC 468

Qy        202 ACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGAAATAT 261
              ||||| ||||                ||||  |||||  || | | | | |  | |  |
Db        469 TCAAATCTAGGGACATACTTTGCGAAAAATACAGGAATTTTAAAGTCTAAAGGAGATATT 528

Qy        262 ATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGA---GCGTCAAATC 318
              ||||  |||   |||  |||||| | || ||| |  ||| | | ||     ||   |
Db        529 ATTTTCTTTCAGGATAGCGATGATGTATGTCACCATGAAAGAATCGAAAGATGTGTTAAT 588

Qy        319 GAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATAAC 378
              | | | |||   ||| ||  |    |  || | | |   |||| |   ||| ||
Db        589 GCATTATTATCGAATAAAGATAATATAGCTGTTAGATGTGCATATTCTAGAATAAATCTA 648

Qy        379 AATAGAAATATTGTTGGCGAAGTTAATGCTCCTCA 413
              | | | | | | |  |||||||| | | |
Db        649 GAAACACAAAATATAATAAAAGTTAATGATAATAA 683
```

RESULT 7
US-11-098-686-9829
; Sequence 9829, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9829
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-9829

Query Match 8.4%; Score 66; DB 17; Length 1146;
Best Local Similarity 49.0%; Pred. No. 0.013;
Matches 205; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

```
Qy         13 ACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAA 72
              ||  |||| ||  |||| || ||||| |||  ||    |    |    |  | |||
Db          7 ACTCTTTCTATTGTTATCCCTGTTTATAATATGGGTTATTGGCTTCCCATTGCTATTGAA 66

Qy         73 TCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACC 132
              ||      ||  |||||  || || |  |||   |  |||  |  | |||  |  |||
Db         67 TCCTGCTTATGGCAATCACATAAAAAAATTGAAGTTATTATTGTAAATGATGGCTCCACA 126

Qy        133 GATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTG 192
              |||  ||||        | | |||  |  ||          |  |||  |  |  | ||
Db        127 GATAATACCCAAGAAGTTATCGAA---CACTATTCTTCTATTACGAATACTATTAGATAT 183

Qy        193 CGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACG 252
              | ||   | ||||   |||    || |          || |   ||||  | | | |||
Db        184 CATACACAACCAAACAAAGGCTCTGGAATTGCCAGACAAGTTGGACAAAACCTTGCTAAA 243

Qy        253 GGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGT 312
              ||  | | |||| | |    ||||||| |||||||| |          |   | |||  |
Db        244 GGTGATTTTATTACATGGATTGATGCTGATGATTTTTTAGCTCCAATGGCAGCAGAACAA 303

Qy        313 CAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATA 372
              |   | || |   || | | ||   |||   |  |||||||||  ||||  |      |
Db        304 CTTTTAAAAATAGCAATTCAAGATAATGTTCCTTTGGTATGTGGTAATGCTGTAACATTT 363

Qy        373 GATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAA 430
                | | || | |  || | | | |    |||| | | | |||  ||  |  ||| ||
Db        364 TCTGATAAAACACTTAATACTAGAGCCAATAATCCACATGATGAAATCCACCATACAA 421
```

RESULT 8
US-11-098-686-8739/c
; Sequence 8739, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8739
; LENGTH: 1457619
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-8739

Query Match 8.4%; Score 66; DB 17; Length 1457619;
Best Local Similarity 49.0%; Pred. No. 0.028;
Matches 205; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

```
Qy         13 ACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAA 72
              ||  |||| ||  |||| || ||||| |||  ||    |    |    |  | |||
Db    1160203 ACTCTTTCTATTGTTATCCCTGTTTATAATATGGGTTATTGGCTTCCCATTGCTATTGAA 1160144
```

```
Qy          73 TCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACC 132
               ||     ||   |||||  || || |  |||   |   |||  |  | ||| |   |||
Db     1160143 TCCTGCTTATGGCAATCACATAAAAAAATTGAAGTTATTATTGTAAATGATGGCTCCACA 1160084

Qy         133 GATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTG 192
               ||| ||||      | | ||| |    |  ||       | |  ||| | |  | ||
Db     1160083 GATAATACCCAAGAAGTTATCGAA---CACTATTCTTCTATTACGAATACTATTAGATAT 1160027

Qy         193 CGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACG 252
               | ||   |  ||||   |||   || |       || |   ||||  | |  | || |
Db     1160026 CATACACAACCAAACAAAGGCTCTGGAATTGCCAGACAAGTTGGACAAAACCTTGCTAAA 1159967

Qy         253 GGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGT 312
               ||  | | |||| | |    ||||||  |||||||| |       |    |  ||| |
Db     1159966 GGTGATTTTATTACATGGATTGATGCTGATGATTTTTTAGCTCCAATGGCAGCAGAACAA 1159907

Qy         313 CAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATA 372
               |   |  || |   || | | ||   |||   | ||||||||  ||||  |       |
Db     1159906 CTTTTAAAAATAGCAATTCAAGATAATGTTCCTTTGGTATGTGGTAATGCTGTAACATTT 1159847

Qy         373 GATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGATAAATTATAGAA 430
               | | || | |  || |  | | |    |||| | | | |||  ||  |  ||| ||
Db     1159846 TCTGATAAAACACTTAATACTAGAGCCAATAATCCACATGATGAAATCCACCATACAA 1159789
```

RESULT 9
US-11-045-004-1
; Sequence 1, Application US/11045004
; Publication No. US20060078901A1
; GENERAL INFORMATION:
; APPLICANT: BUCHRIESER, CARMEN
; APPLICANT: FRANGEUL, LIONEL
; APPLICANT: COUVE, ELISABETH
; APPLICANT: RUSNIOK, CHRISTOPHE
; APPLICANT: FSIHI, HAFIDA
; APPLICANT: DEHOUX, PIERRE
; APPLICANT: DUSSURGET, OLIVIER
; APPLICANT: CHETOUANI, FARID
; APPLICANT: NEDJARI, HAFED
; APPLICANT: GLASER, PHILIPPE
; APPLICANT: KUNST, FRANCK
; APPLICANT: COSSART, PASCALE
; APPLICANT: DANIELS, JUSTIN
; APPLICANT: GOEBEI, WERNER
; APPLICANT: KREFT, JURGEN
; APPLICANT: KUHN, MICHAEL
; APPLICANT: NG, EVA
; APPLICANT: VAZQUEZ-BOLAND, ANTONIO
; APPLICANT: DOMINGUEZ-BERNAL, GUSTAVO
; APPLICANT: GARRIDO-GARCIA, PATRICIA
; APPLICANT: TIERREZ-MARTINEZ, ALBERTO
; APPLICANT: AMEND, ALEXANDRA
; APPLICANT: CHAKRABORTY, TRINAD
; APPLICANT: DOMANN, EUGEN
; APPLICANT: HAIN, THORSTEN
; APPLICANT: BERCHE, PATRICK
; APPLICANT: CHARBIT, ALAIN
; APPLICANT: DURANT, LIONEL
; APPLICANT: PEREZ-DIAZ, JOSE-CLAUDIO
; APPLICANT: BAQUERO, FERNANDO
; APPLICANT: GARCIA DEL PORTILLO, FRANCISCO
; APPLICANT: GOMEZ-LOPEZ, NURIA
; APPLICANT: MADUENIO, ENCARNA
; APPLICANT: PABLOS, BETRIZ DE
; APPLICANT: WEHLAND, JURGEN
; APPLICANT: KARST, UWE
; APPLICANT: ENTIAN, KARL-DIETER
; APPLICANT: HAUF, JORG
; APPLICANT: ROSE, MATTHIAS
; APPLICANT: VOSS, HAMUT
; TITLE OF INVENTION: LISTERIA MONOCYTOGENES GENOME, POLYPEPTIDES AND USES
; FILE REFERENCE: 05394.0018-02
; CURRENT APPLICATION NUMBER: US/11/045,004
; CURRENT FILING DATE: 2005-01-28
; PRIOR APPLICATION NUMBER: 10/637,657
; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: 10/257,023
; PRIOR FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: PCT/FR01/01118
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: FR 00/04,629
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 2854
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 2944528
; TYPE: DNA
; ORGANISM: Listeria monocytogenes
US-11-045-004-1

Query Match 7.5%; Score 59; DB 18; Length 2944528;
Best Local Similarity 49.9%; Pred. No. 0.53;
Matches 183; Conservative 0; Mismatches 175; Indels 9; Gaps 1;

```
Qy          23 TAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCAATTATAC 82
               ||||||| || ||||| |||| |  ||   | | |    ||   | || ||| || |
Db     1124503 TAATTATTCCAGTTTATAATGTAGAAAAGTATGTTAAACGATGCTTGGATTCAGTTCTTG 1124562

Qy          83 ATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGATGATACAT 142
               | ||| | |||||  |  | |  |     |  || | ||| |    || ||| || ||
Db     1124563 AGCAAACATATCATAACCTCGAAGTTATCGTAGTTAATGATGGAGCGACTGATAATTCAG 1124622

Qy         143 TTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGTAACAAGA 202
               | | | || || |||          | |    | ||   ||     |    | | ||
Db     1124623 CTAAAGTCATAAAAAGT---------ATCTCAGATAATCGAATTCGTTATTTTGAAAAAG 1124673

Qy         203 CAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGGAAATATA 262
               ||||   ||    ||       ||||||| |||| |||| |||| |||||  | || |
Db     1124674 AAAATGGCGGACAAGCTACTGCCCGAAATTTTGGATTAGATGTGGCTACGGGAGATTACA 1124733

Qy         263 TTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAAATCGAAG 322
               |      |    |||   |||||||        | | ||   | ||||   |    | ||
Db     1124734 TCGTGATGGTAGATAGTGATGATTACATTAGCAAAAACCTAGTAGAAACTTGTTTGGATA 1124793

Qy         323 TGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGATAACAATA 382
                  || | ||      ||  ||| | || |  | || ||  |||  | |  || |  |
Db     1124794 CTGTACAAAAAACGAATGCTGATTTAGTCTTGTTTACTTCATATAATGTGAATCAGGAAG 1124853

Qy         383 GAAATAT 389
               |||| ||
Db     1124854 GAAAAAT 1124860
```

RESULT 10
US-11-079-463-1654
; Sequence 1654, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAC
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 1654
; LENGTH: 999
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-1654

```
  Query Match             7.3%;  Score 57.4;  DB 18;  Length 999;
  Best Local Similarity   48.8%;  Pred. No. 0.42;
  Matches  185;  Conservative    0;  Mismatches  191;  Indels    3;  Gaps    1;

Qy         16 GTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCA 75
              || ||||| ||  | || ||||| || | |  |       ||| |  ||| |||| |||
Db         31 GTATCAATCATAGTACCGGTTTATAAAGTAGAACGCTACATACACCGTTCAATAGACAGC 90

Qy         76 ATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGAT 135
              |  ||| |||  ||||| |  |||| |    |   |||| |||  |||    | |
Db         91 TTATTACGTCAGACTTATAATAATTTAGAAAT---TATCCTTGTGGATGACGGAAGTCCA 147

Qy        136 GATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGT 195
              ||||  | | |  || ||   |  | |||  ||  |  | |||   |      |
Db        148 GATAATTGTCCTTTCATTTGCGATAAATATGCAATGCAAGATAATCGTATCCGTGTCATC 207

Qy        196 AACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGG 255
              | ||  | |||   ||  |  | ||    || |||  || | ||  |   |  || ||
Db        208 CATAAATCCAATGGCGGCCTCTCTGATGCCCGTAATGCTGCACTAAATGTAATGACAGGA 267

Qy        256 AAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTAGAGCGTCAA 315
              |||||  ||||    |||   |||||  |          || ||  |   | |||
Db        268 CAATATGTCACTTTCGTAGATAGCGATGACTATATCGCAGACAACATGATCGAGAAATTC 327

Qy        316 ATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTATGTTATAGAT 375
              ||||||    ||||      |        || |||||  | ||      || | || |||
Db        328 ATCGAAACAGTAAAACTGTATCAATGTAATATGGTAGTTGCTGGAATGAATATAATTGAT 387

Qy        376 AACAATAGAAATATTGTTG 394
              || ||||   | |||  ||
Db        388 AAAAATAATCAAATTTATG 406


RESULT 11
US-11-194-246-190/c
; Sequence 190, Application US/11194246
; Publication No. US20050272089A1
; GENERAL INFORMATION:
;  APPLICANT: Mott, John
;  APPLICANT:  Trepod, Catherine
;  APPLICANT:  Arvidson, Staffan
;  TITLE OF INVENTION: CRITICAL GENES AND POLYPEPTIDES OF HAEMOPHILUS INFLUENZAE AND MET
;  TITLE OF INVENTION:  USE
;  FILE REFERENCE: 00592.US1 (M&R 268.05920101)
;  CURRENT APPLICATION NUMBER: US/11/194,246
;  CURRENT FILING DATE:  2005-08-01
;  PRIOR APPLICATION NUMBER: US/10/274,586
;  PRIOR FILING DATE: 2002-10-21
;  PRIOR APPLICATION NUMBER: US 60/345,438
;  PRIOR FILING DATE: 2001-10-19
;  NUMBER OF SEQ ID NOS: 621
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 190
;   LENGTH: 2736
;   TYPE: DNA
;   ORGANISM: ARTIFICIAL
;   FEATURE:
;   OTHER INFORMATION: Sequence of the lpp coding sequence and flanking regions.
US-11-194-246-190

  Query Match             7.2%;  Score 56;  DB 17;  Length 2736;
  Best Local Similarity   50.4%;  Pred. No. 0.83;
  Matches  137;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy         16 GTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCA 75
              || ||  | ||| |    | ||| || |  |   ||  ||||       |   |||
Db       2615 GTATCGGTTATTGTTTGTGCTTATAACGCTGAGCAATATATAGATGAAAGCATTTCATCC 2556

Qy         76 ATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGAT 135
              |||||  ||||  ||||| || || | |   |  ||| | ||| ||  || |||
Db       2555 ATTATTAATCAGACTTATGAAAATCTAGAAATTATAGTTATCAATGATGGTTCAACAGAT 2496

Qy        136 GATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGT 195

Db       2495 TTGACTTTGTCTCATTTAGAAGAAATATCTAAATTAGATAAAAGGATAAAAATTATCAGT 2436

Qy        196 AACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGG 255
              || ||    |||||||| |  | | |   |||   ||  | |  |  |||  ||
Db       2435 AATAAATATAATTTAGGGTTCATAAATTCTTTGAATATAGGCCTTGGTTGTTTTTCAGGT 2376

Qy        256 AAATATATTTCTTTTTGTGATGCGGATGATTT 287
              |||||| || |      ||||| |||||| |
Db       2375 AAATATTTTGCAAGAATGGATGCTGATGATAT 2344


RESULT 12
US-11-172-145-3
; Sequence 3, Application US/11172145
; Publication No. US20050272696A1
; GENERAL INFORMATION:
;  APPLICANT: DeAngelis, Paul
;  TITLE OF INVENTION: METHODS OF SELECTIVELY TREATING DISEASES WITH SPECIFIC
;  TITLE OF INVENTION:  GLYCOSAMINOGLYCAN POLYMERS
;  FILE REFERENCE: 3554.104
;  CURRENT APPLICATION NUMBER: US/11/172,145
;  CURRENT FILING DATE:  2005-06-30
;  PRIOR APPLICATION NUMBER: 60/584,442
;  PRIOR FILING DATE: 2004-06-30
;  PRIOR APPLICATION NUMBER: 10/642,248
;  PRIOR FILING DATE: 2003-08-15
;  PRIOR APPLICATION NUMBER: 60/404,356
;  PRIOR FILING DATE: 2002-08-16
;  PRIOR APPLICATION NUMBER: 60/479,432
;  PRIOR FILING DATE: 2003-06-18
;  PRIOR APPLICATION NUMBER: 60/491,362
;  PRIOR FILING DATE: 2003-07-31
;  PRIOR APPLICATION NUMBER: 10/195,908
;  PRIOR FILING DATE: 2002-07-15
;  PRIOR APPLICATION NUMBER: 09/437,277
;  PRIOR FILING DATE: 1999-11-01
;  PRIOR APPLICATION NUMBER: 60/107,929
;  PRIOR FILING DATE: 1998-11-11
;  PRIOR APPLICATION NUMBER: 09/283,402
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/080,414
;  PRIOR FILING DATE: 1998-04-02
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
;   LENGTH: 2979
;   TYPE: DNA
;   ORGANISM: Pasteurella multocida
US-11-172-145-3

  Query Match             7.0%;  Score 55;  DB 17;  Length 2979;
  Best Local Similarity   51.3%;  Pred. No. 1.3;
  Matches  180;  Conservative    0;  Mismatches  165;  Indels    6;  Gaps    2;

Qy         16 GTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGAATCA 75
              ||||| ||   ||| |||| || |  | || ||   |||       ||  |||||
Db       1363 GTTTCTATTTATATCCCCGCTTATAACTGTGCAAATTATATTCAAAGATGTGTAGATAGT 1422

Qy         76 ATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGTAGCACCGAT 135
              | |  ||||| ||  |   ||| | |   | | |||   | ||||| ||   || |||
Db       1423 GCTCTTAATCAAACTGTTGTCGATCTCGAGGTTTGTATTTGTAACGATGGTTCAACAGAT 1482

Qy        136 GATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGAATATTGCGT 195
              ||||  ||    | | |||||  |  |  ||     || ||||  |   || | ||   |
Db       1483 AATACCTTAGAAGTGATCAATAAGCTTTATGGTAATAATCCTAGGGTACGCATCATGTCT 1542

Qy        196 AACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATGGCCACGGGG 255
              ||  |||||   ||  | ||   |      ||||  |   |  |  | ||| |  ||
Db       1543 ---AAACCAAATGGCGGAATAGCCTCAGCATCAAATGCAGCCGTTTCTTTTGCTAAAGGT 1599

Qy        256 AAATATATTTCTTTTTGTGATGCGGATGATT---TGTGGCACGAGAAAAAATTAGAGCGT 312
```

```
Db       1600 TATTACATTGGGCAGTTAGATTCAGATGATTATCTTGAGCCTGATGCAGTTGAACTGTGT 1659

Qy        313 CAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTAT 363
Db       1660 TTAAAAGAATTTTTAAAAGATAAAACGCTAGCTTGTGTTTATACCACTAAT 1710


RESULT 13
US-11-079-463-576
; Sequence 576, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PATH00-03DIV2
;  CURRENT APPLICATION NUMBER: US/11/079,463
;  CURRENT FILING DATE:  2005-03-14
;  PRIOR APPLICATION NUMBER: US 60/128,705
;  PRIOR FILING DATE: 1999-04-09
;  PRIOR APPLICATION NUMBER: US 09/540,209
;  PRIOR FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 576
;   LENGTH: 723
;   TYPE: DNA
;   ORGANISM: B.fragilis
US-11-079-463-576

  Query Match             6.9%;  Score 54.2;  DB 18;  Length 723;
  Best Local Similarity   49.3%;  Pred. No. 1.5;
  Matches  174;  Conservative    0;  Mismatches  173;  Indels    6;  Gaps    1;

Qy          7 AAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCA 66
Db         10 AAAACGATGATAACAGTTTGTATGGCTACCTATAATGGAGAGAAATATATTGAAGAACAA 69

Qy         67 GTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGT 126
Db         70 TTAGAATCAGTACTAATGCAATTGCATTCCAATGATGAAGTTATTATTTCTGATGATGGT 129

Qy        127 AGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGA 186
Db        130 TCTGGCGATAGTACAGTAGACTTGAT------TCGAACCTTCAATGACCCTCGAATAAGG 183

Qy        187 ATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATG 246
Db        184 CTTTTAGTCGGTAATAATTTCTTTTCTCCTACTCAAAACTTTGAAAATGCCTTAAAATAT 243

Qy        247 GCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTA 306
Db        244 GCAAAAGGCGATTATATATTTTTATGTGATCAAGATGATGTTTGGCTTCCCGATAAAGTA 303

Qy        307 GAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAA 359
Db        304 GAAAGTATGCTGCAATATTTGCTTCAATATGATTTGGTTGTGTCGGATTGTAA 356


RESULT 14
US-11-079-463-1041
; Sequence 1041, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PATH00-03DIV2
;  CURRENT APPLICATION NUMBER: US/11/079,463
;  CURRENT FILING DATE:  2005-03-14
;  PRIOR APPLICATION NUMBER: US 60/128,705
;  PRIOR FILING DATE: 1999-04-09
;  PRIOR APPLICATION NUMBER: US 09/540,209
;  PRIOR FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 1041
;   LENGTH: 897
;   TYPE: DNA
;   ORGANISM: B.fragilis
US-11-079-463-1041

  Query Match             6.8%;  Score 53.4;  DB 18;  Length 897;
  Best Local Similarity   47.5%;  Pred. No. 2.1;
  Matches  196;  Conservative    0;  Mismatches  211;  Indels    6;  Gaps    1;

Qy          7 AAAGAAACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCA 66
Db         31 ATAAATATGGTTTCGGTAATAATTCCTTTGTATAATAAGTATTTAGCAATTGGGCGTACT 90

Qy         67 GTAGAATCAATTATACATCAATCTTATCAAGATTTTGTTTTGTATATCATTGACGATTGT 126
Db         91 ATAGAATCAGTAATTGTGCAAACTTATAAAGATTGGGAACTTCTGATTATTGACGATGGT 150

Qy        127 AGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAGA 186
Db        151 TCGAATGACGGCAGTGGGCAAGTGGCTGAGC------AATATACTTTTGATGAGCGTATC 204

Qy        187 ATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGAAATTATGGAATAGAAATG 246
Db        205 CATTATATTTATAAGTCAAATGGTGGTGTATCGTCAGCACGTAATATGGGGATACAAATG 264

Qy        247 GCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAAATTA 306
Db        265 GCGAAAGGTGAATGGCTGCTTTATATTGATGCGGATGATTATCTTTTACCCAATGCTTTA 324

Qy        307 GAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTATTAT 366
Db        325 GAGACCTTATTAAACTTAGCGGAAAAATTTGAAGTAAGTATTGCGGCTAGTAATTTTTAT 384

Qy        367 GTTATAGATAACAATAGAAATATTGTTGGCGAAGTTAATGCTCCTCATGTGAT 419
Db        385 GTTGAGTTTGAAGGAAAAAAGAGGCGTTGCCTTTACAATGTAAGTGAAGGGGT 437


RESULT 15
US-11-045-004-1/c
; Sequence 1, Application US/11045004
; Publication No. US20060078901A1
; GENERAL INFORMATION:
;  APPLICANT: BUCHRIESER, CARMEN
;  APPLICANT:  FRANGEUL, LIONEL
;  APPLICANT:  COUVE, ELISABETH
;  APPLICANT:  RUSNIOK, CHRISTOPHE
;  APPLICANT:  FSIHI, HAFIDA
;  APPLICANT:  DEHOUX, PIERRE
;  APPLICANT:  DUSSURGET, OLIVIER
;  APPLICANT:  CHETOUANI, FARID
;  APPLICANT:  NEDJARI, HAFED
;  APPLICANT:  GLASER, PHILIPPE
;  APPLICANT:  KUNST, FRANCK
;  APPLICANT:  COSSART, PASCALE
;  APPLICANT:  DANIELS, JUSTIN
;  APPLICANT:  GOEBEI, WERNER
;  APPLICANT:  KREFT, JURGEN
;  APPLICANT:  KUHN, MICHAEL
;  APPLICANT:  NG, EVA
;  APPLICANT:  VAZQUEZ-BOLAND, ANTONIO
;  APPLICANT:  DOMINGUEZ-BERNAL, GUSTAVO
;  APPLICANT:  GARRIDO-GARCIA, PATRICIA
;  APPLICANT:  TIERREZ-MARTINEZ, ALBERTO
;  APPLICANT:  AMEND, ALEXANDRA
;  APPLICANT:  CHAKRABORTY, TRINAD
;  APPLICANT:  DOMANN, EUGEN
;  APPLICANT:  HAIN, THORSTEN
;  APPLICANT:  BERCHE, PATRICK
;  APPLICANT:  CHARBIT, ALAIN
```

```
;  APPLICANT:  DURANT, LIONEL
;  APPLICANT:  PEREZ-DIAZ, JOSE-CLAUDIO
;  APPLICANT:  BAQUERO, FERNANDO
;  APPLICANT:  GARCIA DEL PORTILLO, FRANCISCO
;  APPLICANT:  GOMEZ-LOPEZ, NURIA
;  APPLICANT:  MADUENIO, ENCARNA
;  APPLICANT:  PABLOS, BETRIZ DE
;  APPLICANT:  WEHLAND, JURGEN
;  APPLICANT:  KARST, UWE
;  APPLICANT:  ENTIAN, KARL-DIETER
;  APPLICANT:  HAUF, JORG
;  APPLICANT:  ROSE, MATTHIAS
;  APPLICANT:  VOSS, HAMUT
;  TITLE OF INVENTION: LISTERIA MONOCYTOGENES GENOME, POLYPEPTIDES AND USES
;  FILE REFERENCE: 05394.0018-02
;  CURRENT APPLICATION NUMBER: US/11/045,004
;  CURRENT FILING DATE:  2005-01-28
;  PRIOR APPLICATION NUMBER: 10/637,657
;  PRIOR FILING DATE: 2003-08-11
;  PRIOR APPLICATION NUMBER: 10/257,023
;  PRIOR FILING DATE: 2002-10-08
;  PRIOR APPLICATION NUMBER: PCT/FR01/01118
;  PRIOR FILING DATE: 2001-04-11
;  PRIOR APPLICATION NUMBER: FR 00/04,629
;  PRIOR FILING DATE: 2000-04-11
;  NUMBER OF SEQ ID NOS: 2854
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
;   LENGTH: 2944528
;   TYPE: DNA
;   ORGANISM: Listeria monocytogenes
US-11-045-004-1

  Query Match             6.6%;  Score 51.8;  DB 18;  Length 2944528;
  Best Local Similarity   50.1%;  Pred. No. 10;
  Matches  187;  Conservative    0;  Mismatches  177;  Indels    9;  Gaps    2;

Qy         12 AACCGTTTCAATAATTATGCCCGTTTACAATGGGGCCAAAACTATAATCTCATCAGTAGA 71
              ||    | ||  ||||||| ||| |||| |||| ||  ||   ||||     | |  |
Db     534106 AAAAATATCTGTAATTATCCCCATTTATAATGCGGAAAATTATATAGAAAAAGCTTTTTC 534047

Qy         72 ATCAATTATACATCAATCTT------ATCAAGATTTTGTTTTGTATATCATTGACGATTG 125
                | | | ||||| |||       |  | || ||||   | |      |||||||||
Db     534046 CAGTTTAAAAAATCAAACTTTAGCACAAAATGAATTTGAAATTTTATGTATTGACGATAA 533987

Qy        126 TAGCACCGATGATACATTTTCATTAATCAACAGTCGATACAAAAACAATCAGAAAATAAG 185
               ||| ||  | |   ||  |  | | ||     |    ||||| | || | || ||  |
Db     533986 GTCCACAGACAACAGCCTTGGAAAATTAAAACAACTTAGCAAAATCCATAAAAACATGCG 533927

Qy        186 AATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAAAGTCGA---AATTATGGAATAGA 242
              ||| |   |     | | ||||| ||| | | ||   | || |   ||| | |||||  |
Db     533926 GATAATTAATTTGCATAAAAATTCAGGCGGTCCAATGATTCCACGTAATAAAGGAATTAA 533867

Qy        243 AATGGCCACGGGGAAATATATTTCTTTTTGTGATGCGGATGATTTGTGGCACGAGAAAAA 302
              |   || |  || || || |||  ||| |  |||   |||||||       ||  |
Db     533866 AGAAGCAAAAGGAAAGTACATTATTTTCTTAGATAACGATGATTATCTTGGTGAAGAGGC 533807

Qy        303 ATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTAGATGTGGTATGTTCTAATTA 362
              ||| || ||   | |          | ||  |      ||| ||  | |     || ||
Db     533806 ATTGGAACGCTTATTTATTCAAGCTGAAAAAAATCAATCAGAAGTTATTTTCGGGAAATA 533747

Qy        363 TTATGTTATAGAT 375
              |   |  ||| ||
Db     533746 TGTGGGCATAAAT 533734


Search completed: May 11, 2006, 08:30:12
Job time : 413.616 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:45:22 ; Search time 2711.46 Seconds
(without alignments)
15660.244 Million cell updates/sec

Title: US-09-423-093B-2_COPY_2011_2757
Perfect score: 747
Sequence: 1 atgaaaatatcagttataac.....caccatatgataatgaataa 747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	747	100.0	14002	1	AF061251	AF061251 Escherici
	2	747	100.0	14024	6	BD085555	BD085555 Nucleic a
	3	747	100.0	14029	6	AX029262	AX029262 Sequence
c	4	747	100.0	14181	6	BD184789	BD184789 Nucleic a
c	5	747	100.0	14187	6	AR204225	AR204225 Sequence
c	6	747	100.0	14187	6	AR637574	AR637574 Sequence
	7	747	100.0	31482	1	AB008676	AB008676 Escherich
c	8	747	100.0	110000	1	AE005174_28	Continuation (29 o
c	9	747	100.0	110000	1	BA000007_27	Continuation (28 o
	10	343.8	46.0	14323	1	AY730593	AY730593 Salmonell
	11	298.2	39.9	14224	1	AY730592	AY730592 Citrobact
	12	86.8	11.6	2129	1	AF316560	AF316560 Leptospir
	13	86.8	11.6	2129	1	AF316562	AF316562 Leptospir
	14	86.8	11.6	2129	1	AF316563	AF316563 Leptospir
	15	86.8	11.6	2138	1	AF316558	AF316558 Leptospir
	16	86.8	11.6	10029	1	AE011343	AE011343 Leptospir
	17	85.2	11.4	2129	1	AF316557	AF316557 Leptospir
	18	85.2	11.4	36040	1	LIU61226	U61226 Leptospira
c	19	85.2	11.4	302171	1	AE017295	AE017295 Leptospir
	20	83.6	11.2	2129	1	AF316556	AF316556 Leptospir
	21	83.6	11.2	2129	1	AF316564	AF316564 Leptospir
	22	83.6	11.2	2130	1	AF316559	AF316559 Leptospir
	23	83.6	11.2	34723	1	AF316500	AF316500 Leptospir
	24	73.6	9.9	2016	1	AF316561	AF316561 Leptospir
c	25	73.6	9.9	10029	1	AE013290	AE013290 Methanosa
	26	73.4	9.8	10999	1	AE013711	AE013711 Yersinia
	27	73.4	9.8	20588	1	YPS251712	AJ251712 Yersinia
	28	73.4	9.8	20745	1	YPE251713	AJ251713 Yersinia
	29	73.4	9.8	110000	1	BX936398_12	Continuation (13 o
c	30	73.4	9.8	205050	1	AJ414155	AJ414155 Yersinia
	31	73.4	9.8	290510	1	AE017130	AE017130 Yersinia
	32	71.6	9.6	110000	1	BA000022_16	Continuation (17 o
	33	69.2	9.3	110000	1	CR626927_18	Continuation (19 o
	34	69	9.2	3611	1	AY012190	AY012190 Escherich
c	35	68.6	9.2	110000	1	CR626927_30	Continuation (31 o
	36	67.4	9.0	3611	1	AY012189	AY012189 Escherich
c	37	64.4	8.6	110000	1	AY596297_14	Continuation (15 o
c	38	64.2	8.6	178850	8	AL359736	AL359736 Human DNA
c	39	62.8	8.4	110000	1	CP000084_05	Continuation (6 of
	40	62	8.3	302605	1	AE016938	AE016938 Bacteroid
	41	61.8	8.3	3629	1	AY012186	AY012186 Salmonell
	42	61.8	8.3	3629	1	AY012187	AY012187 Salmonell
	43	61.8	8.3	291200	14	AC155129	AC155129 Bos tauru
c	44	60.2	8.1	36233	1	CJ17683CPS	BX545857 Campyloba
	45	58.6	7.8	148387	14	CR932985	CR932985 Danio rer

ALIGNMENTS

RESULT 1
AF061251
LOCUS AF061251 14002 bp DNA linear BCT 20-AUG-1998
DEFINITION Escherichia coli serotype O157:H7 O antigen gene cluster.
ACCESSION AF061251
VERSION AF061251.1 GI:3435170
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 14002)
AUTHORS Wang,L. and Reeves,P.R.
TITLE Organization of Escherichia coli O157 O antigen gene cluster and identification of its specific genes
JOURNAL Infect. Immun. 66 (8), 3545-3551 (1998)
PUBMED 9673232
REFERENCE 2 (bases 1 to 14002)
AUTHORS Wang,L.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1998) Microbiology, The University of Sydney, NSW 2006, Australia
FEATURES Location/Qualifiers
source 1. .14002
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="C664-1992"
/serotype="O157:H7"
/db_xref="taxon:562"
gene 60. .842
/gene="wbdN"
CDS 60. .842
/gene="wbdN"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdN"
/protein_id="AAC32339.1"
/db_xref="GI:3435171"
/translation="MNKETVSIIMPVYNGAKTIISSVESIIHQSYQDFVLYIIDDCST
DDTFSLINSRYKNNQKIRILRNKTNLGVAESRNYGIEMATGKYISFCDADDLWHEKKL
ERQIEVLNNECVDVVCSNYYVIDNNRNIVGEVNAPHVINYRKMLMKNYIGNLTGIYNA

NKLGKFYQKKIGHEDYLMWLEIINKTNGAICIQDNLAYYMRSNNSLSGNKIKAAKWTW
SIYREHLHLSFPKTLYYFLLYASNGVMKKITHSLLRRKETKK"

gene 839. .2023
/gene="wzy"

CDS 839. .2023
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O antigen polymerase Wzy"
/protein_id="AAC32340.1"
/db_xref="GI:3435172"
/translation="MKSAAKLIFLFLFTLYSLQLYGVIIDDRITNFDTKVLTSIIIIF
QIFFVLLFYLTIINERKQQKKFIVNWELKLILVFLFVTIEIAAVVLFLKEGIPIFDDD
PGGAKLRIAEGNGLYIRYIKYFGNIVVFALIILYDEHKFKQRTIIFVYFTTIALFGYR
SELVLLILQYILITNILSKDNRNPKIKRIIGYFLLVGVVCSLFYLSLGQDGEQNDSYN
NMLRIINRLTIEQVEGVPYVVSESIKNDFFPTPELEKELKAIINRIQGIKHQDLFYGE
RLHKQVFGDMGANFLSVTTYGAELLVFFGFLCVFIIPLGIYIPFYLLKRMKKTHSSIN
CAFYSYIIMILLQYLVAGNASAFFFGPFLSVLIMCTPLILLHDTLKRLSRNENISYNC
DL"

gene 1992. .2738
/gene="wbdO"

CDS 1992. .2738
/gene="wbdO"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdO"
/protein_id="AAC32341.1"
/db_xref="GI:3435173"
/translation="MKISVITVTYNNAEGLEKTLSSLSILKIKPFEIIIVDGGSTDGT
NRVISRFTSMNITHVYEKDEGIYDAMNKGRMLAKGDLIHYLNAGDSVIGDIYKNIKEP
CLIKVGLFENDKLLGFSSITHSNTGYCHQGVIFPKNHSEYDLRYKICADYKLIQEVFP
EGLRSLSLITSGYVKYDMGGVSSKKRILRDKELAKIMFEKNKKNLIKFIPISIIKILF
PERLRRVLRKMQYICLTLFFMKNSSPYDNE"

gene 2725. .4116
/gene="wzx"

CDS 2725. .4116
/gene="wzx"
/codon_start=1
/transl_table=11
/product="O antigen flippase Wzx"
/protein_id="AAC32342.1"
/db_xref="GI:3435174"
/translation="MIMNKIKKILKFCTLKKYDTSSALGREQERYRIISLSVISSLIS
KILSLLSLILTVSLTLPYLGQERFGVWMTITSLGAALTFLDLGIGNALTNRIAHSFAC
GKNLKMSRQISGGLTLLAGLSFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKV
FVIIFGLGIYSNGVQKVYMGIQKAYISNIVNAIFILLSIITLVISSKLHAGLPVLIVS
TLGIQYISGIYLTINLIIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSG
DNFIISITLGVTYVAVFSITQRLFQISTVPLTIYNIPLWAAYADAHARNDTQFIKKTL
RTSLKIVGISSFLLAFILVVFGSEVVNIWTEGKIQVPRTFIIAYALWSVIDAFSNTFA
SFLNGLNIVKQQMLAVVTLILIAIPAKYIIVSHFGLTVMLYCFIFIYIVNYFIWYKCS
FKKHIDRQLNIRG"

gene 4113. .5213
/gene="per"

CDS 4113. .5213
/gene="per"
/codon_start=1
/transl_table=11
/product="perosamine synthetase Per"
/protein_id="AAC32343.1"
/db_xref="GI:3435175"
/translation="MKMKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKF
AEQNHVQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFV
DSDNETWQMSVSDIEQKITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEA
FGSKYKGKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHR
QYWHDVIGYNYRMTNICAAIGLAQLEQADDFISRKREIADIYKKNINSLVQVHKESKD
VFHTYWMVSILTRTAEEREELRNHLADKLIETRPVFYPVHTMPMYSEKYQKHPIAEDL
GWRGINLPSFPSLSNEQVIYICESINEFYSDK"

gene 5238. .6452
/gene="wbdP"

CDS 5238. .6452
/gene="wbdP"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdP"
/protein_id="AAC32344.1"
/db_xref="GI:3435176"
/translation="MKIALNSDGFYEWGGGIDFIKYILSILETKPEICIDILLPRNDI
HSLIREKAFPFKSILKAILKRERPRWISLNRFNEQYYRDAFTQNNIETNLTFIKSKSS
AFYSYFDSSDCDVILPCMRVPSGNLNKKAWIGYIYDFQHCYYPSFFSKREIDQRNVFF
KLMLNCANNIIVNAHSVITDANKYVGNYSAKLHSLPFSPCPQLKWFADYSGNIAKYNI
DKDYFIICNQFWKHKDHATAFRAFKIYTEYNPDVYLVCTGATQDYRFPGYFNELMVLA
KKLGIESKIKILGHIPKLEQIELIKNCIAVIQPTLFEGGPGGGVTFDAIALGKKVILS
DIDVNKEVNCGDVYFFQAKNHYSLNDAMVKADESKIFYEPTTLIELGLKRRNACADFL
LDVVKQEIESRS"

gene 6472. .7590
/gene="gmd"

CDS 6472. .7590
/gene="gmd"
/codon_start=1
/transl_table=11
/product="GDP-D-mannose dehydratase Gmd"
/protein_id="AAC32345.1"
/db_xref="GI:3435177"
/translation="MTKVALITGVTGQDGSYLAEFLLDKGYEVHGIKRRASSFNTERI
DHIYQDPHGSNPNFHLHYGDLTDSSNLTRILKEVQPDEVYNLAAMSHVAVSFESPEYT
ADVDAIGTLRLLEAIRFLGLENKTRFYQASTSELYGLVQEIPQKESTPFYPRSPYAVA
KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRGLANIAQGLESCLYL
GNMDSLRDWGHAKDYVRMQWLMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLGIKMSF
VGKGIEEKGIVDSVEGQDAPGVKPGDVIVAVDPRYFRPAEVDTLLGDPSKANLKLGWR
PEITLAEMISEMVAKDLEAAKKHSLLKSHGFSVSLALE"

gene 7587. .8558
/gene="fcl"

CDS 7587. .8558
/gene="fcl"
/codon_start=1
/transl_table=11
/product="fucose synthetase Fcl"
/protein_id="AAC32346.1"
/db_xref="GI:3435178"
/translation="MMMNKQRIFIAGHQGMVGSAITRRLKQRDDVELVLRTRDELNLL
DSSAVLDFFSSQKIDQVYLAAAKVGGILANSSYPADFIYENIMIEANVIHAAHKNNVN
KLLFLGSSCIYPKLAHQPIMEDELLQGKLEPTNEPYAIAKIAGIKLCESYNRQFGRDY
RSVMPTNLYGPNDNFHPSNSHVIPALLRRFHDAVENNSPNVVVWGSGTPKREFLHVDD
MASASIYVMEMPYDIWQKNTKVMLSHINIGTGIDCTICELAETIAKVVGYKGHITFDT
TKPDGAPRKLLDVTLLHQLGWNHKITLHKGLENTYNWFLENQLQYRG"

gene 8561. .9070
/gene="wbdQ"

CDS 8561. .9070
/gene="wbdQ"
/codon_start=1
/transl_table=11
/product="WbdQ"
/protein_id="AAC32347.1"
/db_xref="GI:3435179"
/translation="MFLHSQDFATIVRSTPLISIDLIVENEFGEILLGKRINRPAQGY
WFVPGGRVLKDEKLQTAFERLTEIELGIRLPLSVGKFYGIWQHFYEDNSMGGDFSTHY
IVIAFLLKLQPNILKLPKSQHNAYCWLSRAKLINDDDVHYNCRAYFNNKTNDAIGLDN
KDIICLMRQ"

gene 9052. .10500
/gene="manC"

CDS 9052. .10500
/gene="manC"
/codon_start=1
/transl_table=11
/product="mannose-1-P guanosyltransferase ManC"
/protein_id="AAC32348.1"
/db_xref="GI:3435180"
/translation="MSDAPIIAVVMAGGTGSRLWPLSRELYPKQFLQLSGDNTLLQTT
LLRLSGLSCQKPLVITNEQHRFVVAEQLREINKLNGNIILEPCGRNTAPAIAISAFHA
LKRNPQEDPLLLVLAADHVIAKESVFCDAIKNATPIANQGKIVTFGIIPEYAETGYGY
IERGELSVPLQGHENTGFYYVNKFVEKPNRETAELYMTSGNHYWNSGIFMFKASVYLE
ELRKFRPDIYNVCEQVASSSYIDLDFIRLSKEQFQDCPAESIDFAVMEKTEKCVVCPV
DIGWSDVGSWQSLWDISLKSKTGDVCKGDILTYDTKNNYIYSESALVAAIGIEDMVIV
QTKDAVLVSKKSDVQHVKKIVEMLKLQQRTEYISHREVFRPWGKFDSIDQGERYKVKK
IIVKPGEGLSLRMHHHRSEHWIVLSGTAKVTLGDKTKLVTANESIYIPLGAAYSLENP
GIIPLNLIEVSSGDYLGEDDIIRQKERYKHED"

```
     gene            10504. .11874
                     /gene="manB"
     CDS             10504. .11874
                     /gene="manB"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphomannomutase ManB"
                     /protein_id="AAC32349.1"
                     /db_xref="GI:3435181"
                     /translation="MKSLTCFKAYDIRGKLGEELNEDIAWRIGRAYGEFLKPKTIVLG
                     GDVRLTSEALKLALAKGLQDAGVDVLDIGMSGTEEIYFATFHLGVDGGIEVTASHNPM
                     DYNGMKLVREGARPISGDTGLRDVQRLAEANDFPPVDETKRGRYQQINLRDAYVDHLF
                     GYINVKNLTPLKLVINSGNGAAGPVVDAIEARFKALGAPVELIKVHNTPDGNFPNGIP
                     NPLLPECRDDTRNAVIKHGADMGIAFDGDFDRCFLFDEKGQFIEGYYIVGLLAEAFLE
                     KNPGAKIIHDPRLSWNTVDVVTAAGGTPVMSKTGHAFIKERMRKEDAIYGGEMSAHHY

  Query Match             100.0%;  Score 747;  DB 1;  Length 14002;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-105;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1992 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 2051

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2052 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 2111

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2112 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 2171

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2172 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 2231

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2232 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 2291

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2292 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 2351

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2352 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 2411

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2412 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 2471

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2472 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 2531

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2532 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 2591

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2592 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 2651

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2652 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 2711

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db       2712 AATAGTTCACCATATGATAATGAATAA 2738
```

```
RESULT 2
BD085555
LOCUS       BD085555               14024 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085555
VERSION     BD085555.1  GI:22631165
KEYWORDS    JP 2001522248-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 14024)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 2 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Escherichia coli
            PN   JP 2001522248-A/2
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14024
            FT                   /organism='Escherichia coli'.
FEATURES             Location/Qualifiers
     source          1. .14024
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 747;  DB 6;  Length 14024;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-105;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2011 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 2070

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2071 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 2130

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2131 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 2190

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2191 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 2250

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2251 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 2310

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2311 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 2370

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2371 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 2430

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2431 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 2490

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2491 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 2550

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2551 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 2610

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2611 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 2670

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2671 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 2730

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db       2731 AATAGTTCACCATATGATAATGAATAA 2757
```

RESULT 3
AX029262

```
LOCUS       AX029262               14029 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 2 from Patent WO9850531.
ACCESSION   AX029262
VERSION     AX029262.1  GI:10190113
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 2 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .14029
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
ORIGIN
```

Query Match 100.0%; Score 747; DB 6; Length 14029;
Best Local Similarity 100.0%; Pred. No. 1.6e-105;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy           1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2011 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 2070

Qy          61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2071 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 2130

Qy         121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2131 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 2190

Qy         181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2191 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 2250

Qy         241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2251 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 2310

Qy         301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2311 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 2370

Qy         361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2371 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 2430

Qy         421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2431 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 2490

Qy         481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2491 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 2550

Qy         541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2551 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 2610

Qy         601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2611 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 2670

Qy         661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2671 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 2730

Qy         721 AATAGTTCACCATATGATAATGAATAA 747
               |||||||||||||||||||||||||||
Db        2731 AATAGTTCACCATATGATAATGAATAA 2757
```

RESULT 4
BD184789/c

```
LOCUS       BD184789               14181 bp    DNA     linear   PAT 17-JUN-2003
DEFINITION  Nucleic acid molecule and polypeptide specific to intestinal
            hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
            thereof.
ACCESSION   BD184789
VERSION     BD184789.1  GI:31876989
KEYWORDS    JP 2002355074-A/25.
SOURCE      Escherichia coli O157:H7
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 14181)
  AUTHORS   Hayashi,H., Shinagawa,H., Makino,K., Hayashi,T., Onishi,S.,
            Hattori,M. and Kurokawa,K.
  TITLE     Nucleic acid molecule and polypeptide specific to intestinal
            hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
  JOURNAL   Patent: JP 2002355074-A 25 10-DEC-2002;
            PRESIDENT OF UNIVERSITY OF TSUKUBA
COMMENT     OS   Escherichia coli O157:H7
            PN   JP 2002355074-A/25
            PD   10-DEC-2002
            PF   24-JAN-2002 JP 2002015959
            PI   HIDEO HAYASHI,HIDEO SHINAGAWA,KOZO MAKINO,TETSUYA HAYASHI,SHIN

            PI   ONISHI,
            PI   MASAHIRA HATTORI,KEN KUROKAWA
            PC   C12N15/09,C12N15/09,A61K31/7088,A61K39/00,A61K48/00,A61P31/04,
            PC   C07K14/245,
            PC   C07K16/12,C12M1/00,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/
            PC   02,C12Q1/68,
            PC   G01N33/15,G01N33/50,G01N33/53,G01N33/53,G01N33/566,G01N37/00,
            PC   C12N15/00,
            PC   C12N15/00,C12N5/00
            CC   Nucleic acid molecule and polypeptide specific to intestinal
            CC           hemorrhagic
            CC   pathogenic Escherichia coli O157:H7, and method of use thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14181
            FT                   /organism='Escherichia coli O157:H7'.
FEATURES             Location/Qualifiers
     source          1. .14181
                     /organism="Escherichia coli O157:H7"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:83334"
```

```
ORIGIN

  Query Match             100.0%;  Score 747;  DB 6;  Length 14181;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-105;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11985 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 11926

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11925 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 11866

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11865 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 11806

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11805 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 11746

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11745 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 11686

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11685 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 11626

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11625 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 11566

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11565 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 11506

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11505 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 11446

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11445 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 11386

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11385 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 11326

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11325 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 11266

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db      11265 AATAGTTCACCATATGATAATGAATAA 11239


RESULT 5
AR204225/c
LOCUS       AR204225               14187 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 121 from patent US 6365723.
ACCESSION   AR204225
VERSION     AR204225.1  GI:21500820
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14187)
  AUTHORS   Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
  TITLE     Sequences of E. coli O157
  JOURNAL   Patent: US 6365723-A 121 02-APR-2002;

FEATURES             Location/Qualifiers
     source          1. .14187
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 747;  DB 6;  Length 14187;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-105;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11990 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 11931

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11930 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 11871

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11870 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 11811

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 11751

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 11691

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 11631

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 11571

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 11511

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 11451

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 11391

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 11331

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 11271

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db      11270 AATAGTTCACCATATGATAATGAATAA 11244


RESULT 6
AR637574/c
LOCUS       AR637574               14187 bp    DNA     linear   PAT 20-APR-2005
DEFINITION  Sequence 121 from patent US 6855814.
ACCESSION   AR637574
VERSION     AR637574.1  GI:62771316
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
```

```
REFERENCE   1  (bases 1 to 14187)
  AUTHORS   Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
  TITLE     Sequences of E. coli O157
  JOURNAL   Patent: US 6855814-A 121 15-FEB-2005;
            Wisconsin Alumni Research Foundation; Madison, WI
FEATURES             Location/Qualifiers
     source          1. .14187
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match           100.0%;  Score 747;  DB 6;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-105;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11990 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 11931

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11930 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 11871

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11870 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 11811

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 11751

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 11691

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 11631

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 11571

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 11511

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 11451

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 11391

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 11331

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 11271

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db      11270 AATAGTTCACCATATGATAATGAATAA 11244


RESULT 7
AB008676
LOCUS       AB008676               31482 bp    DNA     linear   BCT 19-MAY-1999
DEFINITION  Escherichia coli 0157 DNA, map position at 46 min., complete cds.
ACCESSION   AB008676
VERSION     AB008676.1  GI:4867913
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Shimizu,T., Yamasaki,S., Tsukamoto,T. and Takeda,Y.
  TITLE     Analysis of the genes responsible for the O-antigen synthesis in
            enterohaemorrhagic Escherichia coli O157
  JOURNAL   Microb. Pathog. 26 (5), 235-247 (1999)
   PUBMED   10222209
REFERENCE   2  (bases 1 to 31482)
  AUTHORS   Shimizu,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1997) Takeshi Shimizu, International Medical
            Center of Japan, Department of Appropriate Technology Development,
            Research Institute; 1-21-1 Toyama, Shinjuku-ku, Tokyo 162, Japan
            (E-mail:kshimizu@imcj.go.jp, Tel:+81-3-3202-7181,
            Fax:+81-3-3202-7364)
FEATURES             Location/Qualifiers
     source          1. .31482
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="184"
                     /serotype="O157"
                     /db_xref="taxon:562"
                     /map="46 min."
                     /clone="1-4-1"
     gene            505. .1398
                     /gene="galF"
     CDS             505. .1398
                     /gene="galF"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="UTP-glucose-1-phosphate uridylyltransferase"
                     /protein_id="BAA77723.1"
                     /db_xref="GI:4867914"
                     /translation="MTNLKAVIPVAGLGMHMLPATKAIPKEMLPIVDKPMIQYIVDEI
                     VAAGIKEILLVTHASKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNV
                     RQGEPLGLGHSILCARPAIGDNPFVVVLPDVVIDDASADPLRYNLAAMIARFNETGRS
                     QVLAKRMPGDLSEYSVIQTKEPLDREGKVSRIVEFIEKPDQPQTLDSDIMAVGRYVLS
                     ADIWPELERTQPGAWGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKY
                     GLRNLKEGAKFRKGIEKLLSE"
     gene            1717. .2499
                     /gene="wbhA"
     CDS             1717. .2499
                     /gene="wbhA"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyl transferase"
                     /protein_id="BAA77724.1"
                     /db_xref="GI:4867915"
                     /translation="MNKETVSIIMPVYNGAKTIISSVESIIHQSYQDFVLYIIDDCST
                     DDTFSLINSRYKNNQKIRILRNKTNLGVAESRNYGIEMATGKYISFCDADDLWHEKKL
                     ERQIEVLNNECVDVVCSNYYVIDNKRNIVGEVNAPHVINYRKMLMKNYIGNLTGIYNA
                     NKLGKFYQKKIGHEDYLMWLEIINKTNGAICIQDNLAYYMRSNNSLSGNKIKAAKWTW
                     SIYREHLHLSFPKTLYYFLLYASNGVMKKITHSLLRRKETKK"
     gene            3153. .3680
                     /gene="wbhB"
     CDS             3153. .3680
                     /gene="wbhB"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase"
                     /protein_id="BAA77725.1"
                     /db_xref="GI:4867916"
                     /translation="MLRIINRLTIEQVESVPYVVSESIKNDFFPTPELEKELKAIINR
                     IQGIKHQDLFYGERLHKQVFGDMGANFLSVTTYGAELLVFFGFLCVFIIPLGIYIPFY
                     LLKRMKKTHSSINCAFYSYIIMILLQYLVAGNASAFFFGPFLSVLIMCTPLILLHDTL
```

```
                KRLSRNENISYNCDL"
gene            3649. .4395
                /gene="wbhC"
CDS             3649. .4395
                /gene="wbhC"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="glycosyl transferase"
                /protein_id="BAA77726.1"
                /db_xref="GI:4867917"
                /translation="MKISVITVTYNNAEGLEKTLSSLSILKIKPFEIIIVDGGSTDGT
                NRVISRFTSMNITHVYEKDEGIYDAMNKGRMLAKGDLIHYLNAGDSVIGDIYKNIKEP
                CLIKVGLFENDKLLGFSSITHSNTGYCHQGVIFPKNHSEYDLRYKICADYKLIQEVFP
                EGLRSLSLITSGYVKYDMGGVSSKKRILRDKELAKIMFEKNKKNLIKFIPISIIKILF
                PERLRRVLRKMQYICLTLFFMKNSSPYDNE"
gene            4388. .5773
                /gene="wzx"
CDS             4388. .5773
                /gene="wzx"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="flippase"
                /protein_id="BAA77727.1"
                /db_xref="GI:4867918"
                /translation="MNKIKKILKFCTLKKYDTSSALGREQERYRIISLSVISSLISKI
                LSLLSLILTVSLTLPYLGQERFGVWMTITSLGAALTFLDLGIGNALTNRIAHSFACGK
                NLKMSRQISGGLTLLAGLSFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKVFV
                IIFGLGIYSNGVQKVYMGIQKAYISNIVNAIFLLLSIITLVISSKLHAGLPVLIVSTL
                GIQYISGIYLTINLIIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSGDN
                FIISITLGVTYVAVFSITQRLFQISTVPLTIYNIPLWAAYADAHARNDTQFIKKTLRT
                SLKIVGISSFLLAFILVVFGSEVVNIWTEGKIQVPRTFIIAYALWSVIDAFSNTFASF
                LNGLNIVKQQMLAVVTLILIAIPAKYIIVSHFGLTVMLYCFIFIYIVNYFIWYKCSFK
                KHIDRQLNIRG"
gene            5776. .6870
                /gene="wbhD"
CDS             5776. .6870
                /gene="wbhD"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="perosamine synthetase"
                /protein_id="BAA77728.1"
                /db_xref="GI:4867919"
                /translation="MKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKFAE
                QNHVQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFVDS
                DNETWQMSVSDIEQKITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEAFG
                SKYKGKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHRQY
                WHDVIGYNYRMTNICAAIGLAQLEQADDFISRKREIADIYKKNINSLVQVHKESKDVF
                HTYWMVSILTRTAEEREELRNHLADKLIETRPVFYPVHTMPMYSEKYQKHPIAEDLGW
                RGINLPSFPSLSNEQVIYICESINEFYSDK"
gene            6895. .8109
                /gene="wbhE"
CDS             6895. .8109
                /gene="wbhE"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="glycosyl transferase"
                /protein_id="BAA77729.1"
                /db_xref="GI:4867920"
                /translation="MKIALNSDGFYEWGGGIDFIKYILSILETKPEICIDILLPRNDI
                HSLIREKAFPFKSILKAILKRERPRWISLNRFNEQYYRDAFTQNNIETNLTFIKSKSS
                AFYSYFDSSDCDVILPCMRVPSGNLNKKAWIGYIYDFQHCYYPSFFSKREIDQRNVFF
                KLMLNCANNIIVNAHSVITDANKYVGNYSAKLHSLPFSPCPQLKWFADYSGNIAKYNI
                DKDYFIICNQFWKHKDHATAFRAFKIYTEYNPDVYLVCTGATQDYRFPGYFNELMVLA
                KKLGIESKIKILGHIPKLEQIELIKNCIAVIQPTLFEGGPGGGVTFDAIALGKKVILS
                DIDVNKEVNCGDVYFFQAKNHYSLNDAMVKADESKIFYEPTTLIELGLKRRNACADFL
                LDVVKQEIESRS"
gene            8129. .9247
                /gene="gmd"
CDS             8129. .9247
                /gene="gmd"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose 4,6-dehydratase"
                /protein_id="BAA77730.1"
                /db_xref="GI:4867921"
                /translation="MTKVALITGVTGQDGSYLAEFLLDKGYEVHGIKRRASSFNTERI
                DHIYQDPHGSNPNFHLHYGDLTDSSNLTRILKEVQPDEVYNLAAMSHVAVSFESPEYT
                ADVDAIGTLRLLEAIRFLGLENKTRFYQASTSELYGLVQEIPQKESTPFYPRSPYAVA
                KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRGLANIAQGLESCLYL
                GNMDSLRDWGHAKDYVRMQWLMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLGIKMSF
                VGKGIEEKGIVDSVEGQDAPGVKPGDVIVAVDPRYFRPAEVDTLLGDPSKANLKLGWR
                PEITLAEMISEMVAKDLEAAKKHSLLKSHGFSVSLALE"
gene            9250. .10215
                /gene="wbhF"
CDS             9250. .10215
                /gene="wbhF"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="GDP-L-fucose pathway enzyme"
                /protein_id="BAA77731.1"
                /db_xref="GI:4867922"
                /translation="MNKQRIFIAGHQGMVGSAITRRLKQRDDVELVLRTRDELNLLDS
                SAVLDFFSSQKIDQVYLAAAKVGGILANSSYPADFIYENIMIEANVIHAAHKNNVNKL
                LFLGSSCIYPKLAHQPIMEDELLQGKLEPTNEPYAIAKIAGIKLCESYNRQFGRDYRS
                VMPTNLYGPNDNFHPSNSHVIPALLRRFHDAVENNSPNVVVWGSGTPKREFLHVDDMA
                SASIYVMEMPYDIWQKNTKVMLSHINIGTGIDCTICELAETIAKVVGYKGHITFDTTK
                PDGAPRKLLDVTLLHQLGWNHKITLHKGLENTYNWFLENQLQYRG"
gene            10218. .10727
                /gene="wbhG"
CDS             10218. .10727
                /gene="wbhG"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="GDP-L-fucose pathway enzyme"
                /protein_id="BAA77732.1"
                /db_xref="GI:4867923"
                /translation="MFLHSQDFATIVRSTPLISIDLIVENEFGEILLGKRINRPAQGY
                WFVPGGRVLKDEKLQTAFERLTEIELGIRLPLSVGKFYGIWQHFYEDNSMGGDFSTHY
                IVIAFLLKLQPNILKLPKSQHNAYCWLSRAKLINDDDVHYNCRAYFNNKTNDAIGLDN
                KDIICLMRQ"
gene            10739. .12157
                /gene="manC"
CDS             10739. .12157
                /gene="manC"
                /note="putative"
                /codon_start=1
                /transl_table=11
```

```
  Query Match             100.0%;  Score 747;  DB 1;  Length 31482;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-105;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3649 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 3708

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3709 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 3768

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3769 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 3828

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3829 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 3888

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
```

```
Db       3889 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 3948

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360

Db       3949 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 4008

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420

Db       4009 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 4068

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480

Db       4069 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 4128

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540

Db       4129 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 4188

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600

Db       4189 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 4248

Qy        601 AATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660

Db       4249 AATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 4308

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720

Db       4309 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 4368

Qy        721 AATAGTTCACCATATGATAATGAATAA 747

Db       4369 AATAGTTCACCATATGATAATGAATAA 4395
```

RESULT 8
AE005174_28/c
WPCOMMENT
Sequence split into 56 fragments LOCUS AE005174 Accession AE005174

Fragment Name	Begin	End
AE005174_00	1	110000
AE005174_01	100001	210000
AE005174_02	200001	310000
AE005174_03	300001	410000
AE005174_04	400001	510000
AE005174_05	500001	610000
AE005174_06	600001	710000
AE005174_07	700001	810000
AE005174_08	800001	910000
AE005174_09	900001	1010000
AE005174_10	1000001	1110000
AE005174_11	1100001	1210000
AE005174_12	1200001	1310000
AE005174_13	1300001	1410000
AE005174_14	1400001	1510000
AE005174_15	1500001	1610000
AE005174_16	1600001	1710000
AE005174_17	1700001	1810000
AE005174_18	1800001	1910000
AE005174_19	1900001	2010000
AE005174_20	2000001	2110000
AE005174_21	2100001	2210000
AE005174_22	2200001	2310000
AE005174_23	2300001	2410000
AE005174_24	2400001	2510000
AE005174_25	2500001	2610000
AE005174_26	2600001	2710000
AE005174_27	2700001	2810000
AE005174_28	2800001	2910000
AE005174_29	2900001	3010000
AE005174_30	3000001	3110000
AE005174_31	3100001	3210000
AE005174_32	3200001	3310000
AE005174_33	3300001	3410000
AE005174_34	3400001	3510000
AE005174_35	3500001	3610000
AE005174_36	3600001	3710000
AE005174_37	3700001	3810000
AE005174_38	3800001	3910000
AE005174_39	3900001	4010000
AE005174_40	4000001	4110000
AE005174_41	4100001	4210000
AE005174_42	4200001	4310000
AE005174_43	4300001	4410000
AE005174_44	4400001	4510000
AE005174_45	4500001	4610000
AE005174_46	4600001	4710000
AE005174_47	4700001	4810000
AE005174_48	4800001	4910000
AE005174_49	4900001	5010000
AE005174_50	5000001	5110000
AE005174_51	5100001	5210000
AE005174_52	5200001	5310000
AE005174_53	5300001	5410000
AE005174_54	5400001	5510000
AE005174_55	5500001	5528445

Continuation (29 of 56) of AE005174 from base 2800001 (AE005174 Escherichia coli O157:H7

Query Match 100.0%; Score 747; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred. No. 7.5e-106;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60

Db      55556 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 55497

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120

Db      55496 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 55437

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180

Db      55436 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 55377

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240

Db      55376 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 55317

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300

Db      55316 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 55257

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360

Db      55256 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 55197

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420

Db      55196 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 55137

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480

Db      55136 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 55077

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540

Db      55076 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 55017

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600

Db      55016 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 54957

Qy        601 AATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660

Db      54956 AATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 54897
```

```
Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54896 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 54837

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db      54836 AATAGTTCACCATATGATAATGAATAA 54810


RESULT 9
BA000007_27/c
WPCOMMENT
  Sequence split into 55 fragments  LOCUS BA000007 Accession BA000007
     Fragment Name      Begin     End
     BA000007_00            1    110000
     BA000007_01       100001    210000
     BA000007_02       200001    310000
     BA000007_03       300001    410000
     BA000007_04       400001    510000
     BA000007_05       500001    610000
     BA000007_06       600001    710000
     BA000007_07       700001    810000
     BA000007_08       800001    910000
     BA000007_09       900001   1010000
     BA000007_10      1000001   1110000
     BA000007_11      1100001   1210000
     BA000007_12      1200001   1310000
     BA000007_13      1300001   1410000
     BA000007_14      1400001   1510000
     BA000007_15      1500001   1610000
     BA000007_16      1600001   1710000
     BA000007_17      1700001   1810000
     BA000007_18      1800001   1910000
     BA000007_19      1900001   2010000
     BA000007_20      2000001   2110000
     BA000007_21      2100001   2210000
     BA000007_22      2200001   2310000
     BA000007_23      2300001   2410000
     BA000007_24      2400001   2510000
     BA000007_25      2500001   2610000
     BA000007_26      2600001   2710000
     BA000007_27      2700001   2810000
     BA000007_28      2800001   2910000
     BA000007_29      2900001   3010000
     BA000007_30      3000001   3110000
     BA000007_31      3100001   3210000
     BA000007_32      3200001   3310000
     BA000007_33      3300001   3410000
     BA000007_34      3400001   3510000
     BA000007_35      3500001   3610000
     BA000007_36      3600001   3710000
     BA000007_37      3700001   3810000
     BA000007_38      3800001   3910000
     BA000007_39      3900001   4010000
     BA000007_40      4000001   4110000
     BA000007_41      4100001   4210000
     BA000007_42      4200001   4310000
     BA000007_43      4300001   4410000
     BA000007_44      4400001   4510000
     BA000007_45      4500001   4610000
     BA000007_46      4600001   4710000
     BA000007_47      4700001   4810000
     BA000007_48      4800001   4910000
     BA000007_49      4900001   5010000
     BA000007_50      5000001   5110000
     BA000007_51      5100001   5210000
     BA000007_52      5200001   5310000
     BA000007_53      5300001   5410000
     BA000007_54      5400001   5498450
Continuation (28 of 55) of BA000007 from base 2700001 (BA000007 Escherichia coli O157:H7

  Query Match             100.0%;  Score 747;  DB 1;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-106;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85342 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 85283

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85282 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 85223

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85222 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 85163

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85162 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 85103

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85102 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 85043

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85042 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 84983

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84982 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 84923

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84922 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 84863

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84862 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 84803

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84802 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 84743

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84742 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 84683

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84682 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 84623

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db      84622 AATAGTTCACCATATGATAATGAATAA 84596


RESULT 10
AY730593
LOCUS       AY730593               14323 bp    DNA     linear   BCT 25-OCT-2004
DEFINITION  Salmonella enterica subsp. enterica serovar Urbana UDP-glucose
            pyrophosphorylase (galF) gene, partial cds; O-antigen gene cluster,
            complete sequence; and 6-phosphogluconate dehydrogenase (gnd) gene,
            partial cds.
ACCESSION   AY730593 AY012194
VERSION     AY730593.1  GI:54401420
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Urbana
  ORGANISM  Salmonella enterica subsp. enterica serovar Urbana
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 14323)
  AUTHORS   Jensen,S.O. and Reeves,P.R.
  TITLE     Molecular evolution of the GDP-mannose pathway genes (manB and
```

manC) in Salmonella enterica
JOURNAL Microbiology (Reading, Engl.) 147 (Pt 3), 599-610 (2001)
PUBMED 11238967
REFERENCE 2 (bases 1 to 14323)
AUTHORS Samuel,G., Hogbin,J.P., Wang,L. and Reeves,P.R.
TITLE Relationships of the Escherichia coli O157, O111, and O55 O-Antigen Gene Clusters with Those of Salmonella enterica and Citrobacter freundii, Which Express Identical O Antigens
JOURNAL J. Bacteriol. 186 (19), 6536-6543 (2004)
PUBMED 15375135
REFERENCE 3 (bases 1 to 14323)
AUTHORS Jenkins,S.O.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-2000) School of Molecular and Microbial Biosciences, University of Sydney, Sydney, NSW 2006, Australia
REFERENCE 4 (bases 1 to 14323)
AUTHORS Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2004) School of Molecular and Microbial Biosciences, University of Sydney, Sydney, NSW 2006, Australia
REMARK Sequence update by submitter
COMMENT On Oct 25, 2004 this sequence version replaced gi:11992707.
FEATURES Location/Qualifiers
source 1. .14323
/organism="Salmonella enterica subsp. enterica serovar Urbana"
/mol_type="genomic DNA"
/strain="M284"
/serotype="30"
/serovar="Urbana"
/sub_species="enterica"
/db_xref="taxon:260677"
gene <1. .852
/gene="galF"
CDS <1. .852
/gene="galF"
/codon_start=1
/transl_table=11
/product="UDP-glucose pyrophosphorylase"
/protein_id="AAV34505.1"
/db_xref="GI:54401421"
/translation="MHMLPATKAIPKEMLPIVDKPMIQYIVDEIVAAGIKEIVLVTHA
SKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNVRQAQPLGLGHSILC
ARPVVGDNPFIVVLPDIIIDDATADPLRYNLAAMVARFNETGRSQVLAKRMKGDLSEY
SVIQTKEPLDNEGKVSRIVEFIEKPDQPQTLDSDLMAVGRYVLSADIWAELERTEPGA
WGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKYGLRNLKEGAKFRKS
IEQLLHE"
gene 1220. .2002
/gene="wbdN"
CDS 1220. .2002
/gene="wbdN"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAV34506.1"
/db_xref="GI:54401422"
/translation="MNREIVSVIMPVYNGSHTIVDSIESVLNQTYKDIKLYVIDDCST
DDTAKLLTVNYANNKNVKIIISSKNKGVAASRNLGIKESKGKYIAFCDADDLWENNKI
EKQINLLEHNNINVVCSNYYIMDFNKKIIGETHSPVTISYKKMLMKNYVGNLTGVYNA
EKLGKFYQKKVGHEDYLMWLEIIKKSNFAICLQENLAFYRVSNNSLSASKLVAAKWTW
NIYRECLGLSLPKSLYYFSLYAANGVIKKIGLINSDNSKAKE"
gene 2255. .3157
/gene="wzy"
CDS 2255. .3157
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O-antigen polymerase"
/protein_id="AAV34507.1"
/db_xref="GI:54401423"
/translation="MEIMAIILFAKEGIPIFDADPGGAKLKIADGNGIYIRYIKYFGN
IVTFTLLLLLNKQKVKQKIIAIIYFLTIALFGYRSELVLLFLQYYIIMSIVNEGGNKY
KFLNKIFIFIFMGGVIGSLFYFSLGQDGAVNNSTENIDRIINRLTIEQVESVPYIISE
SLEYDLFPTPELGKEINAIFNRILGEKDQNLFYGERLHKKVFGDMGANFLSVTTYGAE
LLAFFGPLCIFLIPIGLCIPFAMLKKISNTHSTVNGAIYSFLIMILLQYLIAGNASAF
FFGPFITAIVMSIPLMILHDILKR"
gene 3163. .3894
/gene="wbdO"
CDS 3163. .3894
/gene="wbdO"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAV34508.1"
/db_xref="GI:54401424"
/translation="MNTSVITVTYNNAEGLNKTLESLCLLKVKPYEIIIIDANSIDNT
KAIVKKYSDMNIIYICEKDNGIYDAMNKGLRLARGKLLHYLNAGDCVSGEIYADLNGP
CLLKVGLFDKDNLIGFSSLTHSNTGYCHQGIIFPSKHPIYDLKYKLCADYKLIQMVFP
MGLKQLPLISNGFIKYDMGGVSANKRFLRDKELLQIMLETNKKTLIKFIPLMVIKMVF
PKKIRRILRRVNYYFSLILTNKKNV"
gene 3896. .5287
/gene="wzx"
CDS 3896. .5287
/gene="wzx"
/codon_start=1
/transl_table=11
/product="O-antigen flippase"
/protein_id="AAV34509.1"
/db_xref="GI:54401425"
/translation="MIKEKIKILFKYCTLRKYNTTTDYGRQQERYRIISLSIISNFIS
KLLALLSLILTVSLTLPYLGQDRFGIWMTIASFSAALTFLDLGIGNALTNRIAHSFAR
GKKIRMIRQISGGLTLLAILAFIIIAIFLTIANFVNWGSLLKGINDNTHDEVQFAIKL
FAIIFGIGIFSNGVQRIYMGMQKAYISYISNAFFILISIISLFISSRLHAGIPMLILS
TIGVQYLSGLSLTILLIYKKLITLNKIRINVKKETPYLISNGFFFFILQMGTLATWSG
DNFIISTTLGVTYVSILSITQRLFQISTVPLTIYNAPLWAAYADAHARHDVDFIRKTL
LGSLKIVGISSTTLALILVIFGQSIIAIWTGEELRVPQSLIVAYAVWSIIDAYSNTLS
SFLNGLNIIKQQVFAVVTLILISIPSKYFIVSHFGLTAMLSCFIILYVTNYLIWYKCS
FKKSIHLQLNIRG"
gene 5290. .6384
/gene="per"
CDS 5290. .6384
/gene="per"
/codon_start=1
/transl_table=11
/product="perosamine synthetase"
/protein_id="AAV34510.1"
/db_xref="GI:54401426"
/translation="MEYIPVYQPALSGKEKEYVNDCIDSTWISSKGSYIQRFEKKFAE
QNNIQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPVFVDS
DLQTWQISVDDIKNKITSRTKAIMCVHLYGHPCDMSPILEIAKVNNLFIIEDCAEAFG
SKYKDKFVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAAHRQY
WHDVIGYNYRMTNICAAIGLAQLEQAQLFIARKREIAEIYKNNLTPLIEVHQEGNDIF
HTYWMVSILARNAEEREVLRSFLSEKLTETRPLFYPVHTMPMYSDKYQKHPVAENLGW
RGINLPSFPTLSNEQVIYICNAINEFYKDK"
gene 6409. .7623
/gene="wbdP"
CDS 6409. .7623
/gene="wbdP"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAV34511.1"
/db_xref="GI:54401427"
/translation="MKITLNTDGFNEWGGGVDFIKYLLRTLEIEQDIFLDVLLPREDI
SSLLKRNIFPVKNMIKAILSGQTPRWKKLEGFNEQYYKHAFAEITEQSKLSFITGKSS
AYFSHYIKSDYDIILPCMRTPAVNMKNRAWLGYIYDFQHCYYPSFFSEREIAKRNLFF
KRMLDNANNVIVNSHAVISDAKKFVGDFSANLHALPFSPCPELNWLNDNSDLIDKYAI
DKDYFIICNQFWKHKDHSTAFKAFKKYIEINPDVYLVCTGATQDYRFPGYFDELMCLL
KKLGISEKVKILGHIPKLEQIELIKNSIAVIQPTLFEGGPGGGITFDAVALGKKIVLS
DIEVNKDVDCGHVYFFHSQNEYSLTEALVKATAVDHYYYDPHVLLESGIKRRKKCADFI
LNVISKEIESRI"
gene 7647. .8765
/gene="gmd"
CDS 7647. .8765
/gene="gmd"
/codon_start=1

```
                     /transl_table=11
                     /product="GDP-mannose dehydratase"
                     /protein_id="AAV34512.1"
                     /db_xref="GI:54401428"
                     /translation="MSKVALITGVTGQDGSYLAEFLLEKGYEVHGIKRRASSFNTERV
                     DHIYQDPHSCNPKFHLHYGDLTDASNLTRILQEVQPDEVYNLGAMSHVAVSFESPEYT
                     ADVDAMGTLRLLEAIRFLGLEKKTRFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
                     KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRAIANIAQGLESCLYL
                     GNMDSLRDWGHAKDYVRMQWMMLQQEQPEDFVIATGVQYSVRQFVELAAAQLGIKLRF
                     EGEGINEKGIVVSVTGHDAPGVKPGDVIVAVDPRYFRPAEVETLLGDPSKAHEKLGWK
                     PEITLSEMVSEMVANDLEAAKKHSLLKSHGFNVNLSLE"
     gene            8768. .9733
                     /gene="fcl"
     CDS             8768. .9733
                     /gene="fcl"
                     /codon_start=1
                     /transl_table=11
                     /product="fucose synthetase"
                     /protein_id="AAV34513.1"
                     /db_xref="GI:54401429"
                     /translation="MNKQRIFVAGHRGMVGSAIVRQLAQRGDVELVLRTRDELDLLDG
                     RAVQAFFAGTGIDQVYLAAAKVGGIVANNTYPADFIYENMMIESNIIHAAHLHNVNKL
                     LFLGSSCIYPKLARQPMAESELLQGTLEPTNEPYAIAKIAGIKLCESYNRQYGRDYRS
                     VMPTNLYGPHDNFHPDNSHVIPALLRRFHEAAQSHAPEVVVWGSGTPMREFLHVDDMA
                     AASIHVMELSREVWQENTAPMLSHINVGTGVDCTISELAQTIAKVVGYQGRVVFDAAK
                     PDGTPRKLLDVTRLHQLGWYHEISLEAGLAGTYQWFLENQQRFRG"
     gene            9736. .10236
                     /gene="gmm"
     CDS             9736. .10236
                     /gene="gmm"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-mannose mannosylhydrolase"
                     /protein_id="AAV34514.1"
                     /db_xref="GI:54401430"
                     /translation="MFLRQEDFAAVVRATPLISLDFIVENGQGEILLGQRLNRPAQGY
                     WFVPGGRVRKDETLEAAFARLTQAELGVRLPLAAGTFYGVWQHFYDDNFSGEDFSTHY
                     IVLGFRLRVAESDLLLPDAQHGSYRWLTPEQLLASDNVHDNSRAYFQNAPYSVIGLDK
                     KDVKYV"

  Query Match            46.0%;  Score 343.8;  DB 1;  Length 14323;
  Best Local Similarity  66.3%;  Pred. No. 8.8e-44;
  Matches  495;  Conservative    0;  Mismatches  252;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||| | ||| |||||||| ||||| || |||||||| |||||  || | |||||  |
Db       3163 ATGAATACATCTGTTATAACAGTGACATACAATAATGCAGAAGGACTAAATAAAACACTT 3222

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
                |||||||   |  | ||| | ||||||| ||| || |||||  | ||||      |||
Db       3223 GAGAGTTTATGCTTACTCAAAGTTAAACCTTATGAAATAATTATTATCGATGCAAATTCT 3282

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              |  |||   || ||   | |  | |  | || | |   |||||||| | | || |||
Db       3283 ATTGATAATACTAAGGCTATAGTAAAAAAATATTCGGATATGAATATCATATATATTTGC 3342

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||| ||| | || |||||||||| |||||||| ||||  |   | ||    ||  |
Db       3343 GAAAAGGATAATGGAATATATGATGCAATGAATAAAGGCCTCAGACTAGCACGTGGTAAG 3402

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||| |||||||||||| || || ||  | |||| |||||| || |||     || | ||
Db       3403 TTATTACATTATTTAAATGCTGGTGACTGTGTAAGTGGAGAAATTTATGCCGATTTAAAT 3462

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              |  || |||||  | ||||| || || ||||| || |||||  |  | ||||||||  ||
Db       3463 GGTCCTTGTTTACTAAAAGTAGGTCTGTTCGATAAAGATAATTTAATTGGATTTTCCTCC 3522

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||  |||||||||||  || |||||||| |||||  | |||||  || |  ||  ||| |
Db       3523 TTAACACATTCAAATACTGGATATTGTCACCAAGGTATAATTTTTCCGAGTAAACATCCT 3582

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
                |||||| |||  |||||| ||||||| ||||||||| | || ||   ||| |||||
Db       3583 ATCTATGATTTAAAATATAAACTATGTGCAGATTATAAGTTAATCCAGATGGTTTTTCCA 3642

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              |||||||  |   ||  |||||||| |    || | | | |||||||||||||| ||
Db       3643 ATGGGGTTAAAACAACTTCCTTTGATTTCAAATGGATTTATTAAATATGATATGGGAGGG 3702

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              || ||  | || |||||| |||| ||||||||||| |     |||| |||| || |
Db       3703 GTTTCAGCTAATAAAAGATTTTTGCGAGATAAAGAGTTATTACAAATCATGTTGGAGACT 3762

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||| ||||   | || |||||||||||| |     | |||||||  | |||||  ||
Db       3763 AATAAGAAAACATTAATAAAGTTTATTCCATTAATGGTTATCAAAATGGTTTTCCCGAAA 3822

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
                |||||||| |||| |     |  | |||  || |       | ||    |  ||
Db       3823 AAAATAAGAAGAATATTAAGACGGGTTAATTATTATTTTTCGTTAATTTTAACTAATAAA 3882

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              || |       | ||||| |   | ||
Db       3883 AAAAACGTTTAAAATGATTAAAGAAAA 3909


RESULT 11
AY730592
LOCUS       AY730592               14224 bp    DNA     linear   BCT 02-NOV-2004
DEFINITION  Citrobacter freundii serotype 90 UDP-glucose pyrophosphorylase
            (galF) gene, partial cds; O-antigen gene cluster, complete
            sequence; and 6-phosphogluconate dehydrogenase (gnd) gene, partial
            cds.
ACCESSION   AY730592
VERSION     AY730592.2  GI:55228555
KEYWORDS    .
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1  (bases 1 to 14224)
  AUTHORS   Samuel,G., Hogbin,J.P., Wang,L. and Reeves,P.R.
  TITLE     Relationships of the Escherichia coli O157, O111, and O55 O-Antigen
            Gene Clusters with Those of Salmonella enterica and Citrobacter
            freundii, Which Express Identical O Antigens
  JOURNAL   J. Bacteriol. 186 (19), 6536-6543 (2004)
   PUBMED   15375135
REFERENCE   2  (bases 1 to 14224)
  AUTHORS   Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2004) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
REFERENCE   3  (bases 1 to 14224)
  AUTHORS   Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2004) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
  REMARK    Sequence update by submitter
COMMENT     On Nov 2, 2004 this sequence version replaced gi:54401406.
FEATURES             Location/Qualifiers
     source          1. .14224
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /strain="M1972"
                     /serotype="90"
                     /db_xref="taxon:546"
     gene            <1. .773
                     /gene="galF"
     CDS             <1. .773
                     /gene="galF"
                     /codon_start=3
                     /transl_table=11
                     /product="UDP-glucose pyrophosphorylase"
                     /protein_id="AAV34492.2"
                     /db_xref="GI:55228556"
```

/translation="FVIVAAGIKEIVLVTHASKNAVENHFDTSYELESLLEQRVKRQL
LAEVQSICPPGVTIMNVRQAQLLGLGHSILCARPIIGDNPFVVVLPDVVLDTASADPL
RYNLAAMVARFNETGRSQVLAKRMEGDLSEYSVIQTKDPLDNEGKVSRIVEFIEKPDQ
PQTLDSDLMAVGRYVLSADIWAELEKTAPGAWGRIQLTDAISEPAKKQSVDAILMTGE
SYDCGKKMGYMQAFVKYGLRNLKEGPKFRKSIEKLLAM"

gene 1104. .1853

/gene="wbdN"

CDS 1104. .1853

/gene="wbdN"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAV34493.1"
/db_xref="GI:54401408"
/translation="METESVSIIMPVYNGASTIVQSIESILNQTYKDIRLYIIDDCST
DETSILLQNLYISNDRVELIRNIKNKGVAESRNIGIEQSRGKYIAFCDADDLWGETKL
EKQLDILESEKVDVVCTNYYLIDLEDNIIGSVCSPGIITYKTMLGKNHIGNLTGIYNV
KKIGKVYQKKIGHEDYLMWLEILRKTEQAICIQENLAKYRVSKKSLSANKLIAAQWTW
NIYRKQLKIPIIPALYFFSQYAVQGLLKKII"

gene 1890. .3047

/gene="wzy"

CDS 1890. .3047

/gene="wzy"
/codon_start=1
/transl_table=11
/product="O-antigen polymerase"
/protein_id="AAV34494.1"
/db_xref="GI:54401409"
/translation="MKKSVGIISFFLFSFYVLQLYGLLIDVRIKDVNIQIIAGGLVLF
QILCVLFYALFIRKNLSYQSGLFFNSKKIIIIFLLTFVTIELAAVILFLKEGIPIFDA
DPGGAKLRIADGNGIYIRYIKYFGNIVCFSLCLLLKDRKILHYSVVIIYFSTISLFGY
RSELIMLMIQYYILMSIINNNKVKHSMLRKVVILFSISCIVIGLFYFSLGQEGQENNY
ASNVNRIFNRLTIEQVESVPYIINESMKYDLFPTSEIGKEINAIANRFQGIKDNSLFY
GERLHKDIFGDMGENFLSVTTFGAELLAFFGPFCFFLIPLGIYIPYFLLIKMSQTKNT
VLCSAYSYSVVIFLQYLIAGNFSAFFFGPILTIVIMCIPLVLLHDVLKRLN"

gene 3047. .3781

/gene="wbdO"

CDS 3047. .3781

/gene="wbdO"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAV34495.1"
/db_xref="GI:54401410"
/translation="MKISVVTVTYNNGNGLRKTLQSLSSIKTKPCEIIIIDAESKDNT
KSVVDEYFSKMNIIYINEKDNGIYDAMNKGHKIANGDYIHYLNAGDCVSGDVYANVTG
TCLIKVGLFNERDSFLGFSNLTHSNTGYCHQGVIFPKEHLDYDLKYKLCADYDLLQKT
FPEGLASLPLITDGYIKYDMSGVSANKTLLRDKELLAIMWKNNKRIIPKFMIYVLFKT
MIPKGCRRFLRKLKYSISVFSLNGNK"

gene 3789. .5180

/gene="wzx"

CDS 3789. .5180

/gene="wzx"
/codon_start=1
/transl_table=11
/product="O-antigen flippase"
/protein_id="AAV34496.1"
/db_xref="GI:54401411"
/translation="MNNLLRKIYSVLKYCTLKKFDTSTDYGRQQERYRIISLSIIANL
ISKFLSLLTLVLTVSLTLPYLGPARFGVWMTIASLSAALTFLDLGIGNALTNRIAHAC
ANGERTEVIKKTSGGIALLLLVSICITCVLYIICSVIDWRIFIKGINDNILGEVQRAI
KLFVVIFGISILSNGIQKIYMGMQKAYITYFFNSLLSLISIFCLIYYSSLHATIPVLI
LISSGIPLFTGMLLMILLIKDGFVSLNEIFHNMRNEAPHLISTGFFFFILQIGTLATW
SGDNFIISSTLGVAFVAILSLTQRLFQISTVPLTIYNTPLWVAYADAHARKDTHFIKK
TLLASFKIVGFSSLIMATFLILFGRDIISVWTGDKIDIPQSFIIAYAIWSIIDAVSNT
LASFLNGLGIVKQQMFAVVSLVIISIPAKYFIVSHFGLTAMLSCFIAIYIVNYIVWYT
ISFKKCIYSQLSI"

gene 5188. .6285

/gene="per"

CDS 5188. .6285

/gene="per"
/codon_start=1
/transl_table=11
/product="perosamine synthetase"
/protein_id="AAV34497.1"
/db_xref="GI:54401412"
/translation="MKYIPVYEPSLTGKEKEYVNDCLDSTWISSKGSYIQRFETEFSK
HISAQYATTVSNGTVAIHLALIALGISEGDEVIVPTLTYIASVNAINYTGAIPVFVDS
DAETWQVSIEDIRNKITNKTKAIMCVHLYGQPCDMDAILEIANEFNLFVIEDCAEAFG
SKYKDKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYERCVHYKGQGLAAHRQY
WHDVIGYNYRMTNICAAIGLAQLEQAEGFINRKRNIAEIYKKELFNSPVKVHQESDNT
FHTYWMVSILTRTAEEREELRNYLSENFIETRPLFYPVHTMPMYSEKYQKHPVAEDLG
WRGMNLPSYPGMTDEQVIYICNKIKCFYNEK"

gene 6310. .7524

/gene="wbdP"

CDS 6310. .7524

/gene="wbdP"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAV34498.2"
/db_xref="GI:55228557"
/translation="MKITINTDGFNEWGGGVDFIKYLLLILSTKDDVFTDIIVPKDDF
ASLFRRNVFPIKNAIKSLGKGKFPRWVKKDGFDEGYYHNAFSGLVDKSKLTFINSNSN
SYFDYHAKSDNDIIFPCMRVPTNSVCDSPWIGYIYDFQHCYYPSFFSAREIKKRELYF
REMLSSANNVIVNANSVIADAKNFIGDFSADLHALPFSPCPQQIWLDDNSNLVDKYAI
SKDYFIVCNQFWKHKNHVVVFKAFKEYIKVNPEAYLVCTGGIQDYRFPEYFGELKALI
QKLGISSKIKILGHIPKLEQVELIKKSIAVIQPTLFEGGPGGGATYDAVALGKKVILS
DIDVNREVNCGDVIFFNPQDEYSLAKALFKAEEEKSVKENGDLMKLGLKRREKCADFI
MSVIRKEIESRN"

gene 7546. .8664

/gene="gmd"

CDS 7546. .8664

/gene="gmd"
/codon_start=1
/transl_table=11
/product="GDP-mannose dehydratase"
/protein_id="AAV34499.1"
/db_xref="GI:54401414"
/translation="MSKVALITGVTGQDGSYLAEFLLEKGYEVHGIKRRASSFNTGRV
DHIYQDPHLGNPKFNLHYGDLTDTSNLTRILKEVQPDEVYNLGAMSHVAVSFESPEYT
ADVDAMGTLRLLEAIRFLGLEKKTRFYQASTSELYGLVREIPQKETTPFYPRSPYAVA
KLYAYWITVNYRESYGMYACNGILFNHESPRRGETFVTRKITRAMANIAQGLESCLYL
GNMDSLRDWGHAKDYVRMQWMMLQQEQPEDFVIATGVQYSVRDFVEMSAAQLGIKLSF
EGDGVNEKGIVASVTGHDAPGVKVGDVIVEVDPRYFRPAEVETLLGDPSKAHEKLGWK
PEITLQEMVAEMVASDLEAAKKYSLLKESGFDVNLSLE"

gene 8667. .9632

/gene="fcl"

CDS 8667. .9632

/gene="fcl"
/codon_start=1
/transl_table=11
/product="fucose synthetase"
/protein_id="AAV34500.1"
/db_xref="GI:54401415"
/translation="MKKQRIFVAGHRGMVGTAIVRQLSLRDNVELVLRTRDELNLLDS
SAVQAFFATERIDQVYLAAAKVGGIVANNTYPADFIYENMMIESNIIHAAHLHNVNKL
LFLGSSCIYPKQATQPIAESELLQGTLEPTNEPYAIAKIAGIKLCESYNRQYGRDYRS
VMPTNLYGPHDNFHPSNSHVIPALLRRFHEAREQNAPDVVVWGSGTPMREFLHVDDMA
AASIHVMELDSEIWQEYTQPMLSHINVGTGVDCTIRELAQTIAQVVGYKGKVVFDSSK
PDGTPRKLLDVTRLHSLGWRHSVSLEYGLESTYQWFLENQHWCRG"

gene 9635. .10135

/gene="gmm"

CDS 9635. .10135

/gene="gmm"
/codon_start=1
/transl_table=11
/product="GDP-mannose mannosylhydrolase"
/protein_id="AAV34501.1"
/db_xref="GI:54401416"
/translation="MFLHHDEFATVVRSTPLVSIDLIVENADGEFLLGKRLNRPAQGY
WFVPGGRVQKDERIENAFERLTLAELGLRLPLSVAEFYGVWQHFYDDNFSGPDFSTHY
IVLAFHIRGQESCLCLPDEQHKIWRWFKSEELEEEKDVHVNSRAYFHIENSNVIGLDN
KGLNHV"

gene 10128. .11576

/gene="manC"

```
     CDS             10128. .11576
                     /gene="manC"
                     /codon_start=1
                     /transl_table=11
                     /product="mannose-1-P guanosyltransferase"
                     /protein_id="AAV34502.1"
                     /db_xref="GI:54401417"

  Query Match           39.9%;  Score 298.2;  DB 1;  Length 14224;
  Best Local Similarity  64.8%;  Pred. No. 8.5e-37;
  Matches  476;  Conservative    0;  Mismatches  253;  Indels    6;  Gaps    2;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              |||||||| ||||| |||||||| |||||||| |||| | | ||||||  ||| ||||||
Db       3047 ATGAAAATTTCAGTAGTAACTGTAACTTATAACAATGGTAATGGGTTAAGAAAGACTTTA 3106

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
                ||||||||   | | ||'|  ||||| | ||| || |||||  | ||||  |  ||
Db       3107 CAAAGTTTATCTTCTATTAAAACTAAACCATGTGAAATAATTATTATAGATGCAGAGTCA 3166

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGA---TTTACTAGTATGAATATTACACATGTT 177
              |  |||   || ||    ||  |   |  |   ||| |||  ||||||||||   || |
Db       3167 AAGGATAATACTAAGTCAGTTGTAGATGAATATTTTTCTAAAATGAATATTATCTATATA 3226

Qy        178 TATGAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGC 237
               | ||||||||| | || || || ||||| |||||||| || |  |   | || || ||
Db       3227 AACGAAAAAGATAATGGTATTTACGATGCAATGAATAAAGGGCACAAAATTGCGAACGGT 3286

Qy        238 GACTTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATC 297
              || |  || |||||| | || || || ||| | ||    || ||| | ||   ||| |
Db       3287 GATTACATCCATTATCTTAATGCAGGTGATTGTGTTTCAGGCGATGTTTACGCAAACGTA 3346

Qy        298 AAAGAGCCATGTTTGATTAAAGTTGGCCTTTTC---GAAAATGATAAACTTCTGGGATTT 354
              | || | |||||||| ||||||||||  | |||   || |  |||    || |||| ||
Db       3347 ACAGGGACATGTTTAATTAAAGTTGGATTGTTCAATGAGAGAGATTCCTTTTTGGGGTTC 3406

Qy        355 TCTTCTATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAAT 414
              |||    |||| || |||||||||| ||||| ||||| || || ||||| |||||  |
Db       3407 TCTAACTTAACGCACTCAAATACAGGTTATTGCCATCAGGGTGTTATTTTTCCAAAAGAA 3466

Qy        415 CATTCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTG 474
              || | ||| |||||||| | ||||||| | |||||||||| |  |  | ||| ||
Db       3467 CACTTAGATTATGATCTGAAGTATAAACTTTGTGCTGATTATGATTTATTGCAAAAGACT 3526

Qy        475 TTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATG 534
              ||||| || ||||||   || ||  | || || |||   || ||  | || |||||||||
Db       3527 TTTCCGGAGGGGTTAGCCTCCCTTCCATTAATAACTGATGGATACATTAAGTATGATATG 3586

Qy        535 GGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTT 594
              | ||||| ||| |||| |||| | |  || | ||||||||  |       |||||||
Db       3587 AGCGGAGTCTCTGCAAATAAAACATTACTACGTGATAAAGAATTGCTTGCTATTATGTGG 3646

Qy        595 GAAAAAAATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTC 654
              |||| ||||||   |  ||  ||||||||      |   || | ||||   |  |
Db       3647 AAAAACAATAAACGGATTATTCCAAAGTTTATGATTTATGTTTTATTTAAAACCATGATT 3706

Qy        655 CCTGAACGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTC 714
              ||| || |||   || |  | ||| | ||| || |||||  |  |  |  ||| |   ||
Db       3707 CCTAAAGGTTGTCGACGTTTTTTGAGAAAATTGAAATATTCTATTTCAGTTTTTTCACTC 3766

Qy        715 ATGAAGAATAGTTCA 729
              |    ||||  | |
Db       3767 AATGGAAATAAATGA 3781


RESULT 12
AF316560
LOCUS       AF316560               2129 bp    DNA     linear   BCT 20-DEC-2001
DEFINITION  Leptospira interrogans strain Lai isolate LAI2122 rfb locus,
            partial sequence.
ACCESSION   AF316560
VERSION     AF316560.1  GI:17940057
KEYWORDS    .
SOURCE      Leptospira interrogans
  ORGANISM  Leptospira interrogans
            Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE   1  (bases 1 to 2129)
  AUTHORS   de la Pena-Moctezuma,A., Bulach,D.M. and Adler,B.
  TITLE     Genetic differences among the LPS biosynthetic loci of serovars of
            Leptospira interrogans and Leptospira borgpetersenii
  JOURNAL   FEMS Immunol. Med. Microbiol. 31 (1), 73-81 (2001)
   PUBMED   11476985
REFERENCE   2  (bases 1 to 2129)
  AUTHORS   de la Pena-Moctezuma,A., Bulach,D.M. and Adler,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2000) Department of Microbiology, Monash
            University, P.O. Box 53, Monash University, VIC 3800, Australia
FEATURES             Location/Qualifiers
     source          1. .2129
                     /organism="Leptospira interrogans"
                     /mol_type="genomic DNA"
                     /strain="Lai"
                     /serovar="Lai"
                     /isolate="LAI2122"
                     /db_xref="taxon:173"
     CDS             1. .798
                     /note="OrfL21"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAL49454.1"
                     /db_xref="GI:17940058"
                     /translation="MTAKEPKISIITINLNNLEGLRKTLESVKSQTYTNFELIVVDGG
                     STDGSFEYLKSNLDLIKKFISEKDKGIYNAQNKGILLSKGEYLVFLNAGDTLLQKNIL
                     LEISKFLDQDVDLVYGNILIDSKDNGIIERKYPDRLNYFYWSIKSLCHQAVFIRKSLF
                     DLYGYYNEEYLFAADFEFFHRFWFNKNIKIKHASVFVTLYDFNGVSAQPKNRKRIAEE
                     YQKIKKKYFPIWAYVVYKVNSYLLERFSNTFVWNLLFKFYRFINPKR"
     CDS             1308. .2129
                     /note="OrfL22"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAL49455.1"
                     /db_xref="GI:17940059"
                     /translation="MYNTPIVSVIIPCYNYGKYIEQAIQSILEQSYKNWEIIVVDDGS
                     DDEYTIEKLEELKKEYAVIKIDRSGPAVARNVGIEAAKGKFILPLDSDDTIHSDYLLE
                     AISAYEKKPSLGIVYCEAEFFGSMKGRWNLPKYSFPEILLDNCIFVSAVFRKSDWKEV
                     GGFNGNMKNEWEDYDFWLSLIEKGRDVYKIPRVMFYYRRGHVSRSSRSIETYLPLYLQ
                     LFKNHKRLYTENIEVLFKRHLKARELEEQFSILTKNPIIYGIVRFLVSCLKFFAK"
ORIGIN

  Query Match           11.6%;  Score 86.8;  DB 1;  Length 2129;
  Best Local Similarity  52.8%;  Pred. No. 0.0004;
  Matches  187;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy          4 AAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGT 63
              ||||| ||| | || ||| | | ||  ||||||   |||||||||   |||||||||
Db         19 AAAATTTCAATAATTACTATTAATTTAAATAATCTGGAAGGGTTACGTAAAACTTTAGAA 78

Qy         64 AGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACA 123
              ||| |          |     | |  ||||||  | ||  | || ||||| || || ||
Db         79 AGTGTTAAGTCGCAGACTTATACAAATTTTGAATTGATAGTCGTAGATGGAGGTTCAACG 138

Qy        124 GATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAA 183
              |||||||    |       | |    |         | | |  |  |    | | | ||||
Db        139 GATGGAAGTTTTGAATATTTAAAATCAAACTTAGATTTAATCAAAAAGTTTATCTCTGAA 198

Qy        184 AAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTA 243
              |||||| ||||||| ||| ||||    || || ||     || |  | |||||| || |
Db        199 AAAGATAAAGGGATTTATAATGCACAAAACAAAGGAATTTTGCTTTCTAAAGGAGAGTAT 258

Qy        244 ATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAAGAG 303
                |   || |||||| || || ||||   |  |   | | |    |  |      |
Db        259 CTCGTTTTTTTAAATGCGGGAGATACTTTGCTACAAAAAAATATTCTATTAGAGATTTCT 318

Qy        304 CCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCT 357
```

```
Db        319 AAATTTTTAGACCAAGATGTTGATTTAGTATATGGAAATATACTGATAGATTCT 372


RESULT 13
AF316562
LOCUS       AF316562                2129 bp    DNA     linear   BCT 20-DEC-2001
DEFINITION  Leptospira interrogans strain Naam isolate NAA2122 rfb locus,
            partial sequence.
ACCESSION   AF316562
VERSION     AF316562.1  GI:17940063
KEYWORDS    .
SOURCE      Leptospira interrogans
  ORGANISM  Leptospira interrogans
            Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE   1  (bases 1 to 2129)
  AUTHORS   de la Pena-Moctezuma,A., Bulach,D.M. and Adler,B.
  TITLE     Genetic differences among the LPS biosynthetic loci of serovars of
            Leptospira interrogans and Leptospira borgpetersenii
  JOURNAL   FEMS Immunol. Med. Microbiol. 31 (1), 73-81 (2001)
  PUBMED    11476985
REFERENCE   2  (bases 1 to 2129)
  AUTHORS   de la Pena-Moctezuma,A., Bulach,D.M. and Adler,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2000) Department of Microbiology, Monash
            University, P.O. Box 53, Monash University, VIC 3800, Australia
FEATURES             Location/Qualifiers
     source          1. .2129
                     /organism="Leptospira interrogans"
                     /mol_type="genomic DNA"
                     /strain="Naam"
                     /serovar="Naam"
                     /isolate="NAA2122"
                     /db_xref="taxon:173"
     CDS             1. .798
                     /note="OrfN21"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAL49458.1"
                     /db_xref="GI:17940064"
                     /translation="MIAKEPKISIITINLNNLEGLRKTLESVKSQTYTNFELIVVDGG
                     STDGSFEYLKSNLDLIKKFISEKDKGIYNAQNKGILLSKGEYLVFLNAGDTLLQKNIL
                     LEISKFLDQDVDLVYGNILIDSKDNGIIERKYPDRLNYFYWSIKSLCHQAVFIRKSLF
                     DLYGYYNEEYLFAADFEFFHRFWFNKNIKIKHASVFVTLYDFNGVSAQPKNRKRIAEE
                     YQKIKKKYFPIWAYVVYKVNSYLLERFSNTFVWNLLFKFYRFINPKR"
     CDS             1308. .2129
                     /note="OrfN22"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAL49459.1"
                     /db_xref="GI:17940065"
                     /translation="MYNTPIVSVIIPCYNYGKYIEQAIQSILEQSYKNWEIIVVDDGS
                     DDEYTIEKLEELKKKYAVIKIDRSGPAVARNVGIEAAKGKFILPLDSDDTIHSDYLLE
                     AISAYEKKPSLGIVYCEAEFFGSMKGRWNLPKYSFPEILLDNCIFVSAVFRKSDWKEV
                     GGFNGNMKNEWEDYDFWLSLIEKGRDVYKIPRVMFYYRRGHVSRSSRSIETYLPLYLQ
                     LFKNHKRLYTENIEVLFKRHLKARELEEQFSILTKNPIIYGIVRFLVSCLKFFAK"
ORIGIN

  Query Match           11.6%;  Score 86.8;  DB 1;  Length 2129;
  Best Local Similarity  52.8%;  Pred. No. 0.0004;
  Matches  187;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy          4 AAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGT 63
              ||||| ||| | || ||| | | ||  ||||||   |||||||||   |||||||||
Db         19 AAAATTTCAATAATTACTATTAATTTAAATAATCTGGAAGGGTTACGTAAAACTTTAGAA 78

Qy         64 AGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACA 123
              ||| |          |     | |  ||||||  | ||  | || ||||| || || ||
Db         79 AGTGTTAAGTCGCAGACTTATACAAATTTTGAATTGATAGTCGTAGATGGAGGTTCAACG 138

Qy        124 GATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAA 183
              |||||||     |       | |     |       | | |  |  |    | | | ||||
Db        139 GATGGAAGTTTTGAATATTTAAAATCAAACTTAGATTTAATCAAAAAGTTTATCTCTGAA 198

Qy        184 AAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTA 243
              |||||| |||||||  ||| ||||    || || ||      || |  | ||||| || |
Db        199 AAAGATAAAGGGATTTATAATGCACAAAACAAAGGAATTTTGCTTTCTAAAGGAGAGTAT 258

Qy        244 ATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAAGAG 303
              |   || |||||| || || ||||    |  |   | | |    |  |       |
Db        259 CTCGTTTTTTTAAATGCGGGAGATACTTTGCTACAAAAAAATATTCTATTAGAGATTTCT 318

Qy        304 CCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCT 357
              || |||     ||| ||    ||| | | |||  ||  | |||  |  ||||
Db        319 AAATTTTTAGACCAAGATGTTGATTTAGTATATGGAAATATACTGATAGATTCT 372


RESULT 14
AF316563
LOCUS       AF316563                2129 bp    DNA     linear   BCT 20-DEC-2001
DEFINITION  Leptospira interrogans strain Salinem isolate PYR2122 rfb locus,
            partial sequence.
ACCESSION   AF316563
VERSION     AF316563.1  GI:17940066
KEYWORDS    .
SOURCE      Leptospira interrogans
  ORGANISM  Leptospira interrogans
            Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE   1  (bases 1 to 2129)
  AUTHORS   de la Pena-Moctezuma,A., Bulach,D.M. and Adler,B.
  TITLE     Genetic differences among the LPS biosynthetic loci of serovars of
            Leptospira interrogans and Leptospira borgpetersenii
  JOURNAL   FEMS Immunol. Med. Microbiol. 31 (1), 73-81 (2001)
  PUBMED    11476985
REFERENCE   2  (bases 1 to 2129)
  AUTHORS   de la Pena-Moctezuma,A., Bulach,D.M. and Adler,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2000) Department of Microbiology, Monash
            University, P.O. Box 53, Monash University, VIC 3800, Australia
FEATURES             Location/Qualifiers
     source          1. .2129
                     /organism="Leptospira interrogans"
                     /mol_type="genomic DNA"
                     /strain="Salinem"
                     /serovar="Pyrogenes"
                     /isolate="PYR2122"
                     /db_xref="taxon:173"
     CDS             1. .798
                     /note="OrfY21"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAL49460.1"
                     /db_xref="GI:17940067"
                     /translation="MIAKEPKISIITINLNNLEGLRKTLESVKSQTYTNFELIVVDGG
                     STDGSFEYLKSNLDLIKKFISEKDKGIYNAQNKGILLSKGEYLVFLNAGDTLLQKNIL
                     LEISKFLDQDVDLVYGNILIDSKDNGIIERKYPDRLNYFYWSIKSLCHQAVFIRKSLF
                     DLYGYYNEEYLFAADFEFFHRFWFNKNIKIKHASVFVTLYDFNGVSAQPKNRKRIAEE
                     YQKIKKKYFPIWAYVVYKVNSYLLERFSNTFVWNLLFKFYRFINPKR"
     CDS             1308. .2129
                     /note="OrfY22"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAL49461.1"
                     /db_xref="GI:17940068"
                     /translation="MYNTPIVSVIIPCYNYGKYIEQAIQSILEQSYKNWEIIVVDDGS
                     DDEYTIEKLEELKKKYAVIKIDRSGPAVARNVGIEAAKGKFILPLDSDDTIHSDYLLE
                     AIAAYEKKPSLGIVYCEAEFFGSMKGKWNLPKYSFPEILLDNCIFVSAVFRKSDWKEV
                     GGFNGNMKNEWEDYDFWLSLIEKGRDVYKIPRVMFYYRRGHVSRSSRSIETYLPLYLQ
                     LFKNHKRLYTENIEVLFKRHLKARELEEQFSILTKNPIIYGIVRFLVSCLKFFAK"
ORIGIN

  Query Match           11.6%;  Score 86.8;  DB 1;  Length 2129;
```

```
  Best Local Similarity   52.8%;  Pred. No. 0.0004;
  Matches  187;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy          4 AAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGT 63
              ||||| ||| | || ||| | | ||  ||||||   |||||||||   |||||||||
Db         19 AAAATTTCAATAATTACTATTAATTTAAATAATCTGGAAGGGTTACGTAAAACTTTAGAA 78

Qy         64 AGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACA 123
              ||| |          |      | |  ||||||  | ||  | || ||||| || || ||
Db         79 AGTGTTAAGTCGCAGACTTATACAAATTTTGAATTGATAGTCGTAGATGGAGGTTCAACG 138

Qy        124 GATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAA 183
              |||||||     |        | |    |        | | |  |  |    | | | ||||
Db        139 GATGGAAGTTTTGAATATTTAAAATCAAACTTAGATTTAATCAAAAAGTTTATCTCTGAA 198

Qy        184 AAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTA 243
              |||||| ||||||| ||| ||||    || || ||      || |  | ||||| || |
Db        199 AAAGATAAAGGGATTTATAATGCACAAAACAAAGGAATTTTGCTTTCTAAAGGAGAGTAT 258

Qy        244 ATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAAGAG 303
              |   || |||||| || || ||||   |  |   | | |    |  |      |
Db        259 CTCGTTTTTTTAAATGCGGGAGATACTTTGCTACAAAAAAATATTCTATTAGAGATTTCT 318

Qy        304 CCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCT 357
                || |||     ||| ||    ||| | | |||  ||  | |||  |  ||||
Db        319 AAATTTTTAGACCAAGATGTTGATTTAGTATATGGAAATATACTGATAGATTCT 372


RESULT 15
AF316558
LOCUS       AF316558               2138 bp    DNA     linear   BCT 20-DEC-2001
DEFINITION  Leptospira interrogans strain Swart isolate BAT2122 rfb locus,
            partial sequence.
ACCESSION   AF316558
VERSION     AF316558.1  GI:17940051
KEYWORDS    .
SOURCE      Leptospira interrogans
  ORGANISM  Leptospira interrogans
            Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE   1  (bases 1 to 2138)
  AUTHORS   de la Pena-Moctezuma,A., Bulach,D.M. and Adler,B.
  TITLE     Genetic differences among the LPS biosynthetic loci of serovars of
            Leptospira interrogans and Leptospira borgpetersenii
  JOURNAL   FEMS Immunol. Med. Microbiol. 31 (1), 73-81 (2001)
   PUBMED   11476985
REFERENCE   2  (bases 1 to 2138)
  AUTHORS   de la Pena-Moctezuma,A., Bulach,D.M. and Adler,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2000) Department of Microbiology, Monash
            University, P.O. Box 53, Monash University, VIC 3800, Australia
FEATURES             Location/Qualifiers
     source          1. .2138
                     /organism="Leptospira interrogans"
                     /mol_type="genomic DNA"
                     /strain="Swart"
                     /serovar="Bataviae"
                     /isolate="BAT2122"
                     /db_xref="taxon:173"
     CDS             1. .798
                     /note="OrfB21"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAL49450.1"
                     /db_xref="GI:17940052"
                     /translation="MIAKEPKISIITINLNNLEGLRKTLESVKSQTYTNFELIVVDGG
                     STDGSFEYLKSNLDLIKKFISEKDKGIYNAQNKGILLSKGEYLVFLNAGDTLLQKNIL
                     LEISKFLDQDVDLVYGNILIDSKDNGIIERKYPDRLNYFYWSIKSLCHQAVFIRKSLF
                     DLYGYYNEEYLFAADFEFFHRFWFNKNIKIKHASVFVTLYDFNGVSAQPKNRKRIAEE
                     YQKIKKKYFPIWAYVVYKVNSYLLERFSNTFVWNLLFKFYRFINPKR"
     CDS             1317. .2138
                     /note="OrfB22"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAL49451.1"
                     /db_xref="GI:17940053"
                     /translation="MYNTPIVSVIIPCYNYGKYIEQAIQSILEQSYKNWEIIVVDDGS
                     DDEYTIEKLEELKKKYAVIKIDRSGPAVARNVGIEAAKGKFILPLDSDDTIHSDYLLE
                     AISAYEKKPSLGIVYCEAEFFGSMKGRWNLPKYSFPEILLDNCIFVSAVFRKSDWKEV
                     GGFNGNMKNEWEDYDFWLSLIEKGRDVYKIPRVMFYYRRGHVSRSSRSIETYLPLYLQ
                     LFKNHKRLYTENIEVLFKRHLKARELEEQFSILTKNPIIYGIVRFLVSCLKFFAK"
ORIGIN

  Query Match             11.6%;  Score 86.8;  DB 1;  Length 2138;
  Best Local Similarity   52.8%;  Pred. No. 0.0004;
  Matches  187;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy          4 AAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGT 63
              ||||| ||| | || ||| | | ||  ||||||   |||||||||   |||||||||
Db         19 AAAATTTCAATAATTACTATCAATTTAAATAATCTGGAAGGGTTACGTAAAACTTTAGAA 78

Qy         64 AGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACA 123
              ||| |          |      | |  ||||||  | ||  | || ||||| || || ||
Db         79 AGTGTTAAGTCGCAGACTTATACAAATTTTGAATTGATAGTCGTAGATGGAGGTTCAACG 138

Qy        124 GATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAA 183
              |||||||     |        | |    |        | | |  |  |    | | | ||||
Db        139 GATGGAAGTTTTGAATATTTAAAATCAAACTTAGATTTAATCAAAAAGTTTATCTCTGAA 198

Qy        184 AAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTA 243
              |||||| ||||||| ||| ||||    || || ||      || |  | ||||| || |
Db        199 AAAGATAAAGGGATTTATAATGCACAAAACAAAGGAATTTTGCTTTCTAAAGGAGAGTAT 258

Qy        244 ATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAAGAG 303
              |   || |||||| || || ||||   |  |   | | |    |  |      |
Db        259 CTCGTTTTTTTAAATGCGGGAGATACTTTGCTACAAAAAAATATTCTATTAGAGATTTCT 318

Qy        304 CCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCT 357
                || |||     ||| ||    ||| | | |||  ||  | |||  |  ||||
Db        319 AAATTTTTAGACCAAGATGTTGATTTAGTATATGGAAATATACTGATAGATTCT 372


Search completed: May 10, 2006, 20:33:41
Job time : 2714.46 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:42:47 ; Search time 348.71 Seconds
(without alignments)
14276.972 Million cell updates/sec

Title: US-09-423-093B-2_COPY_2011_2757
Perfect score: 747
Sequence: 1 atgaaaatatcagttataac.....caccatatgataatgaataa 747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	747	100.0	14024	2	AAX06749	Aax06749 E. coli O
	2	747	100.0	14024	3	AAZ56386	Aaz56386 Escherich
c	3	747	100.0	14181	10	ADC01214	Adc01214 Enterohae
c	4	747	100.0	14187	9	ACD19108	Acd19108 E. coli 0
	5	83.2	11.1	14309	14	AEC31432	Aec31432 E. coli t
	6	77.4	10.4	14657	14	AEA31026	Aea31026 Shigella
	7	73.4	9.8	744	8	ACA54039	Aca54039 Prokaryot
c	8	63.6	8.5	2000	11	ACL37108	Acl37108 Rice stre
	9	62.8	8.4	13749	14	AEA29447	Aea29447 Escherich
c	10	56.4	7.6	1024	6	ABX66419	Abx66419 Helicobac
	11	55.8	7.5	9415	14	AEA39955	Aea39955 Escherich
	12	55.8	7.5	15390	14	AEA33353	Aea33353 E. coli O
	13	55.2	7.4	14516	2	AAX06748	Aax06748 E. coli O
	14	55.2	7.4	14516	3	AAZ56375	Aaz56375 Escherich
	15	53.2	7.1	10432	14	AEA31080	Aea31080 Shigella
c	16	51.8	6.9	50000	6	ABL55643	Abl55643 AmEPV gen
	17	51.6	6.9	1011	8	ACA27781	Aca27781 Prokaryot
c	18	51.4	6.9	4985	6	ABQ75107	Abq75107 Anopheles
c	19	51.4	6.9	4985	10	ACF79720	Acf79720 Mosquito
c	20	51	6.8	29993	10	ADB37663	Adb37663 Human che
c	21	51	6.8	29993	10	ADB37661	Adb37661 Human che
c	22	51	6.8	38342	4	AAS46746	Aas46746 Tumour su
c	23	51	6.8	38342	6	ABK31507	Abk31507 Signal tr
	24	50.8	6.8	1367	14	ADZ70967	Adz70967 Human chr
c	25	50.8	6.8	5647	6	ABL33566	Abl33566 Human imm
c	26	50.8	6.8	5647	6	ABL70355	Abl70355 Chemicall
c	27	50.8	6.8	5647	6	AAS61320	Aas61320 Human gen
	28	50.8	6.8	8056	8	ABZ10246	Abz10246 Haematopo
c	29	50.6	6.8	10189	6	ABL34161	Abl34161 Human imm
	30	50.6	6.8	158001	12	ADL17884	Adl17884 Human pho
c	31	50.4	6.7	8056	8	ABZ10246	Abz10246 Haematopo
c	32	49.6	6.6	3315	14	ADW71859	Adw71859 OKT3Vk an
	33	49.6	6.6	6992	3	AAZ60930	Aaz60930 Nucleotid
	34	49.4	6.6	26281	3	AAZ60929	Aaz60929 Nucleotid
c	35	49	6.6	647	13	ADR59482	Adr59482 Cotton cD
	36	49	6.6	8056	8	ABZ10100	Abz10100 Haematopo
c	37	48.8	6.5	2000	11	ACL35363	Acl35363 Rice stre
c	38	48.8	6.5	34722	13	ABD33520	Abd33520 Human can
	39	48.6	6.5	333	5	ABV12865	Abv12865 Human pro
c	40	48.4	6.5	1393	1	AAN60225	Aan60225 Sequence
	41	48.4	6.5	2000	11	ACL37108	Acl37108 Rice stre
c	42	48.2	6.5	8170	6	ABK28258	Abk28258 DNA trans
c	43	48	6.4	2987	13	ADS89604	Ads89604 Oligonucl
c	44	48	6.4	2987	13	ADS89330	Ads89330 Oligonucl
c	45	48	6.4	5987	6	ABL33563	Abl33563 Human imm

ALIGNMENTS

RESULT 1
AAX06749
ID AAX06749 standard; DNA; 14024 BP.
XX
AC AAX06749;
XX
DT 26-APR-1999 (first entry)
XX
DE E. coli O157 antigen gene cluster.
XX
KW O antigen; O157 antigen; diagnosis; wbdN gene; wzx gene; wzy gene;
KW wbdP gene; wbdO gene; wbdR gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Escherichia coli.
XX

FH	Key	Location/Qualifiers
FT	CDS	79. .861
FT		/*tag= a
FT		/product= "sugar_transferase"
FT		/note= "wbdN gene (ORF1), this region is specifically
FT		claimed in Claim 6; encodes AAW88312"
FT	CDS	858. .2042
FT		/*tag= b
FT		/product= "O_antigen-polymerase"
FT		/note= "wzy gene (ORF2), this region is specifically
FT		claimed in Claim 6; encodes AAW88313"
FT	CDS	2011. .2757
FT		/*tag= c
FT		/product= "sugar_transferase"
FT		/note= "wbdO gene (ORF3), this region is specifically
FT		claimed in Claim 6; encodes AAW88314"
FT	CDS	2744. .4135
FT		/*tag= d
FT		/product= "O_antigen_flippase"
FT		/note= "wzx gene (ORF4), this region is specifically
FT		claimed in Claim 6; encodes AAW88315"
FT	CDS	4132. .5232
FT		/*tag= e
FT		/product= "perosamine_synthetase"
FT		/note= "per gene (ORF5); encodes AAW88316"
FT	CDS	5257. .6471

```
FT                   /*tag=  f
FT                   /product= "sugar_transferase"
FT                   /note= "wbdP gene (ORF6), this region is specifically
FT                   claimed in Claim 6; encodes AAW88317"
FT   CDS             6491. .7609
FT                   /*tag=  g
FT                   /note= "ORF7; encodes AAW88318"
FT   CDS             7606. .8578
FT                   /*tag=  h
FT                   /note= "ORF8; encodes AAW88319"
FT   CDS             8580. .9089
FT                   /*tag=  i
FT                   /note= "ORF9; encodes AAW88320"
FT   CDS             9071. .10519
FT                   /*tag=  j
FT                   /note= "manC gene (ORF10); encodes AAW88321"
FT   CDS             10523. .11893
FT                   /*tag=  k
FT                   /note= "manD gene (ORF11); encodes AAW88322"
FT   repeat_region   12007. .13123
FT                   /tag= l
FT                   /note= "H-repeat, may play a role in gene cluster
FT                   assembly"
FT   CDS             13156. .13821
FT                   /*tag=  m
FT                   /product= "N-acetyl_transferase"
FT                   /note= "wbdR gene (ORF12), this region is specifically
FT                   claimed in Claim 6; encodes AAW88323"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;  98WO-AU000315.
XX
PR   01-MAY-1997;  97AU-00006545.
PR   22-JUL-1997;  97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88312, AAW88313, AAW88314, AAW88315, AAW88316, AAW88317,
DR   AAW88318, AAW88319, AAW88320, AAW88321, AAW88322, AAW88323.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 6; Fig 8; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Escherichia coli O157 O antigen. The gene cluster was
CC   obtained by PCR amplification of E. coli O157:H7 strain (C664-1992) DNA
CC   using primers (see AAX06752-63) based on O antigen gene sequences. 12
CC   Open reading frames within the gene cluster were identified on the basis
CC   of homology to known sequences. The genes encode proteins (see AAW88312-
CC   23) involved the synthesis of sugars present in the polysaccharide
CC   antigen, and in the transport or processing of polysaccharide or
CC   oligosaccharide units. The use of nucleic acid molecules derived from
CC   particular assembly and transport genes, particularly wbd (transferase),
CC   wzx (flippase) and wzy (polymerase) genes, within O antigen gene clusters
CC   improves the specificity of methods for the detection and identification
CC   of O antigens, e.g. in tests of food- or faecal- derived samples, or
CC   samples from patients. The O antigen is a major virulence factor of
CC   enteropathogenic E. coli strains that cause diarrhoea and haemorrhagic
CC   colitis
XX
SQ   Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 747;  DB 2;  Length 14024;
  Best Local Similarity 100.0%;  Pred. No. 5.1e-142;

  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2011 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 2070

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2071 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 2130

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2131 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 2190

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2191 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 2250

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2251 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 2310

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2311 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 2370

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2371 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 2430

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2431 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 2490

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2491 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 2550

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2551 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 2610

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2611 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 2670

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2671 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 2730

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db       2731 AATAGTTCACCATATGATAATGAATAA 2757


RESULT 2
AAZ56386
ID   AAZ56386 standard; DNA; 14024 BP.
XX
AC   AAZ56386;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:56.
XX
KW   Flagellin; fliC; antigen; detection; ds.
XX
OS   Escherichia coli.
XX
PN   WO9961458-A1.
XX
PD   02-DEC-1999.
XX
```

PF 21-MAY-1999; 99WO-AU000385.
XX
PR 21-MAY-1998; 98AU-00003634.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Reeves PR, Wang L;
XX
DR WPI; 2000-072598/06.
XX
PT Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, feces, etc.
XX
PS Claim 3; Page 229-233; 245pp; English.
XX
CC AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all or
CC part of an Escherichia coli flagellin protein except a protein expressed
CC by E. coli H1, H7, H12 or H48 type strains. The present invention also
CC describes a method of detecting the presence of E. coli of a particular H
CC serotype in a sample, comprising specifically hybridising a nucleic acid,
CC preferably at least a pair, derived from a flagellating gene, specific
CC for a particular flagellin gene associated with the H serotype, to any
CC E.coli in the sample which contain the gene, and detecting any hybridised
CC molecules, identifying the presence of that serotype in the sample. (I)
CC are useful for: (1) detecting the presence of E. coli of H serotype in a
CC sample by hybridising at least one or a pair of (I) to any E. coli in the
CC sample and detecting the hybridised nucleic acid molecules; and (2) for
CC detecting the presence of both O and H-serotypes of E. coli by
CC hybridising at least one or a pair of (I) to any E. coli present in the
CC sample and detecting the hybridised nucleic acid molecules. (I) is
CC particularly useful for detecting the combination of O and H antigen.
CC Hybridised (I) when using at least one (I) is detected by southern blot
CC analysis and, when using a pair of (I), is detected by polymerase chain
CC reaction (PCR). AAZ56399 to AAZ56420 represent primers used in the
CC exemplification of the present invention
XX
SQ Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 U; 0 Other;

Query Match 100.0%; Score 747; DB 3; Length 14024;
Best Local Similarity 100.0%; Pred. No. 5.1e-142;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2011 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 2070

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2071 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 2130

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2131 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 2190

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2191 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 2250

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2251 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 2310

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2311 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 2370

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2371 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 2430

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2431 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 2490

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2491 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 2550

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2551 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 2610

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2611 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 2670

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2671 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 2730

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db       2731 AATAGTTCACCATATGATAATGAATAA 2757
```

RESULT 3
ADC01214/c
ID ADC01214 standard; DNA; 14181 BP.
XX
AC ADC01214;
XX
DT 04-DEC-2003 (first entry)
XX
DE Enterohaemorragic E. coli 0157:H7-specific nucleic acid SEQ ID NO: 1258.
XX
KW ds; gene; enterohaemorragic; anti-bacterial.
XX
OS Escherichia coli; 0157:H7.
XX
PN JP2002355074-A.
XX
PD 10-DEC-2002.
XX
PF 24-JAN-2002; 2002JP-00015959.
XX
PR 24-JAN-2001; 2001JP-00112010.
XX
PA (UYTS-) UNIV TSUKUBA.
XX
DR WPI; 2003-451640/43.
XX
PT Enterohemorragic Escherichia coli O157:H7-specific nucleic acid molecule
PT and a polypeptide and its use, a polypeptide, a vector and a host cell.
XX
PS Claim 2; SEQ ID NO 1258; 2067pp; Japanese.
XX
CC The invention relates to a novel enterohaemorragic Escherichia coli
CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
CC has anti-bacterial activity. The polypeptide can be used in detection
CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
CC genome of Enterohaemorragic E coli O157:H7 was determined. The present
CC sequence represents an E. coli 0157:H7-specific nucleic acid of the
CC invention.
XX
SQ Sequence 14181 BP; 4450 A; 2856 C; 2281 G; 4594 T; 0 U; 0 Other;

Query Match 100.0%; Score 747; DB 10; Length 14181;
Best Local Similarity 100.0%; Pred. No. 5.1e-142;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11985 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 11926

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      11925 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 11866

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11865 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 11806

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11805 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 11746

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11745 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 11686

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11685 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 11626

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11625 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 11566

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11565 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 11506

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11505 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 11446

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11445 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 11386

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11385 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 11326

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11325 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 11266

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db      11265 AATAGTTCACCATATGATAATGAATAA 11239
```

```
RESULT 4
ACD19108/c
ID   ACD19108 standard; DNA; 14187 BP.
XX
AC   ACD19108;
XX
DT   27-OCT-2003  (revised)
DT   21-AUG-2003  (first entry)
XX
DE   E. coli 0157 unique DNA sequence OZID_121.
XX
KW   OZID; ds; acute haemorrhagic colitis; haemolytic uraemic syndrome;
KW   food poisoning.
XX
OS   Escherichia coli; strain 0157:H7.
XX
PN   US2003023075-A1.
XX
PD   30-JAN-2003.
XX
PF   01-APR-2002; 2002US-00114170.
XX
PR   04-DEC-1998;   98US-0110955P.
PR   03-DEC-1999;   99US-00453702.
XX
PA   (BLAT/) BLATTNER F R.
PA   (BURL/) BURLAND V D.
PA   (PERN/) PERNA N T.
PA   (PLUN/) PLUNKETT G.
PA   (WELC/) WELCH R.
XX
PI   Blattner FR,  Burland VD,  Perna NT,  Plunkett G,  Welch R;
XX
DR   WPI; 2003-479497/45.
XX
PT   New DNA sequences from Escherichia coli strain O157:H7, useful for
PT   detecting E. coli O157:H7 in a sample, or in designing diagnostic probes
PT   which can be used to distinguish strain O157:H7 from strain K12 using
PT   molecular techniques.
XX
PS   Claim 16; SEQ ID NO 121; 33pp; English.
XX
CC   The invention relates to an isolated DNA molecule comprising an E. coli
CC   strain 0157:H7 sequence selected from a clostridial cytotoxin-like gene,
CC   a urease gene cluster, a RTX toxin-like gene cluster, a locus of
CC   enterocyte effacement and 2 genes from its associated lymphocytic phage
CC   933W (a putative serine/threonine kinase and a tail fibre gene). E. coli
CC   0157:H7 can cause food poisoning, specifically acute haemorrhagic colitis
CC   (which can develop into haemolytic uraemic syndrome). Also included are
CC   an isolated DNA molecule comprising a nucleotide sequence identical to at
CC   least 25 contiguous nucleotides contained in DNA sequences selected from
CC   ACD18988-ACD19242 (being 255 E.coli 0157 DNA sequences which are not
CC   found in E.coli K12), a recombinant DNA construction comprising the DNA
CC   above and a method for detecting E. coli 0157:H7 (ATCC 43895) in a sample
CC   (or distinguishing between 0157 and K12) using a probe derived from one
CC   of the 255 sequences. The DNA sequences are useful in detecting E. coli
CC   O157:H7 in a sample, for the early diagnosis of humans and livestock
CC   infected with O157:H7, and in designing diagnostic probes which can be
CC   used to distinguish strain O157:H7 from strain K12 using molecular
CC   techniques. The present sequence is one of the 255 E. coli 0157:H7 DNA
CC   sequence (termed OZID_1-OZID255). Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from the USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030023075 (Updated on 27-OCT-2003
CC   to standardise OS field)
XX
SQ   Sequence 14187 BP; 4451 A; 2859 C; 2282 G; 4595 T; 0 U; 0 Other;
```

```
  Query Match           100.0%;  Score 747;  DB 9;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 5.1e-142;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11990 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 11931

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11930 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 11871

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11870 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 11811

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 11751

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 11691

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 11631

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 11571
```

```
Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 11511

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 11451

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 11391

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 11331

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 11271

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              ||||||||||||||||||||||||||
Db      11270 AATAGTTCACCATATGATAATGAATAA 11244
```

RESULT 5
AEC31432
ID AEC31432 standard; DNA; 14309 BP.
XX
AC AEC31432;
XX
DT 20-OCT-2005 (first entry)
XX
DE E. coli type O43 O-antigen gene cluster.
XX
KW O-antigen; gene expression; DNA detection; gene; ds.
XX
OS Escherichia coli.
XX

FH	Key	Location/Qualifiers
FT	CDS	1083. .2396
FT		/*tag= a
FT		/note= "Orf1"
FT	CDS	2412. .3155
FT		/*tag= b
FT		/note= "Orf2. Specifically claimed in Claim 3"
FT	CDS	3152. .3673
FT		/*tag= c
FT		/note= "Orf3. Specifically claimed in Claim 3"
FT	CDS	3652. .4740
FT		/*tag= d
FT		/note= "Orf4. Specifically claimed in Claim 3"
FT	CDS	4712. .5938
FT		/*tag= e
FT		/note= "Orf5"
FT	CDS	5935. .6984
FT		/*tag= f
FT		/note= "Orf6. Specifically claimed in Claim 3"
FT	CDS	6981. .8099
FT		/*tag= g
FT		/note= "Orf7"
FT	CDS	8103. .9068
FT		/*tag= h
FT		/note= "Orf8"
FT	CDS	9071. .9526
FT		/*tag= i
FT		/note= "Orf9"
FT	CDS	9538. .10941
FT		/*tag= j
FT		/note= "Orf10"
FT	CDS	11007. .11753
FT		/*tag= k
FT		/note= "Orf11. Specifically claimed in Claim 3"
FT	CDS	11778. .13148
FT		/*tag= l
FT		/note= "Orf12"

XX
PN CN1563043-A.
XX
PD 12-JAN-2005.
XX
PF 19-APR-2004; 2004CN-00019028.
XX
PR 19-APR-2004; 2004CN-00019028.
XX
PA (TIAN-) TIANJIN BIOCHIP TECH CO LTD.
XX
PI Wang L, Yang J, Feng L;
XX
DR WPI; 2005-307118/32.
XX
PT Nucleotide peculiar to O-antigen of Escherichia coli O43.
XX
PS Claim 1; SEQ ID NO 1; 35pp; Chinese.
XX
CC The invention relates to a nucleotide sequence (SEQ ID No:1) which is
CC specific to the O-antigen of Escherichia coli O43. The nucleotide
CC sequence represents the complete sequence (14309bp) of a gene cluster for
CC controlling the synthesis of O-antigen in Escherichia coli O43. The
CC nucleotide sequence can have one or several inserted, deleted or
CC substituted bases, whilst still retaining its function. Also described is
CC a PCR primer for the glycosyltransferase (galF) gene and a PCR primer for
CC the oligosaccharide unit treatment (gnd) gene, both of which are
CC contained within the E. coli type O43 O-antigen gene cluster. The
CC invention also discloses a method for detecting and identifying E. coli
CC type O43 in humans. This sequence represents the E. coli type O43 O-
CC antigen gene cluster of the invention.
XX
SQ Sequence 14309 BP; 4529 A; 2267 C; 2972 G; 4541 T; 0 U; 0 Other;

Query Match 11.1%; Score 83.2; DB 14; Length 14309;
Best Local Similarity 57.2%; Pred. No. 2.2e-07;
Matches 151; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

```
Qy          4 AAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGT 63
              ||| ||||| | | |||||    ||||||  |          | |||||  |  | | |
Db       2421 AAATTATCAATAGTGACTGTCGTATATAATGGGGAGAGGTTCATTGAAAATGCAATCAAT 2480

Qy         64 AGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACA 123
              ||| |   |     |||||   | |   ||||   |||||||||| ||||| || ||||||
Db       2481 AGTGTTATATCCCAAAAAAATGAGCAAGTTGAATATATTATAGTAGATGGGGGGTCTACA 2540

Qy        124 GATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAA 183
              |||  ||||     | | ||    | || ||  | ||  |    ||    | |||  |||
Db       2541 GATTCAACGTTAAATATTATAGAAAAATATAAAAATAAAATAGATATTTTTATTTCAGAA 2600

Qy        184 AAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTA 243
              |||||| |||| |||||||||| |||||||| ||  | |     || |  || || |
Db       2601 AAAGATAAAGGTATATATGATGCAATGAATAAAGGATGGAGCAAAGCTACTGGAGAATAT 2660

Qy        244 ATACATTATTTAAACGCCGGCGAT 267
              |   ||  || ||    |  |||
Db       2661 GTGGCTTTCTTGAATAGTGATGAT 2684
```

RESULT 6
AEA31026
ID AEA31026 standard; DNA; 14657 BP.
XX
AC AEA31026;
XX
DT 28-JUL-2005 (first entry)
XX
DE Shigella dysenteriae 4 O-antigen gene cluster DNA.
XX

KW microorganism detection; O-antigen; operon; ds.
XX
OS Shigella dysenteriae.
XX
PN CN1563067-A.
XX
PD 12-JAN-2005.
XX
PF 19-APR-2004; 2004CN-00019052.
XX
PR 19-APR-2004; 2004CN-00019052.
XX
PA (TIAN-) TIANJIN BIOCHIP TECH CO LTD.
XX
PI Wang L, Kong Q, Feng L;
XX
DR WPI; 2005-307142/32.
XX
PT Nucleotide peculiar to O-antigen of Shigella dysenteriae type 4 useful
PT for controlling synthesis of O-antigen and detecting and identifying
PT Shigella dysenteriae type 4 in human body.
XX
PS Claim 1; SEQ ID NO 1; 35pp; Chinese.
XX
CC The invention relates to a novel Shigella dysenteriae 4 O-antigen
CC polynucleotide comprising the full-length nucleotide sequence of the gene
CC cluster responsible for controlling O-antigen synthesis in Shigella
CC dysenteriae 4. The cluster includes the glycosyl transferase genes and
CC oligosaccharide unit treatment genes. The polynucleotide of the invention
CC may be used for detecting the O-antigen and thus the Shigella dysenteriae
CC 4 bacterium in the human body or in the environment, using methods such
CC as PCR. The current sequence is that of the Shigella dysenteriae 4 O-
CC antigen gene cluster DNA of the invention which contains the galF, wzx,
CC orf2, orf3, orf4, wzy, orf6, gmd, fcl, gmm, manC, orf11, manB and gnd
CC genes.
XX
SQ Sequence 14657 BP; 4504 A; 2410 C; 3015 G; 4728 T; 0 U; 0 Other;

Query Match 10.4%; Score 77.4; DB 14; Length 14657;
Best Local Similarity 56.3%; Pred. No. 3.3e-06;
Matches 166; Conservative 0; Mismatches 126; Indels 3; Gaps 1;

```
Qy          4 AAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGT 63
              |||||  | |||||||||| | | ||| |||||||| ||||| |||  ||||||  |   |
Db       2934 AAAATTACCGTTATAACTATAAATTACAATAATGCAGAAGGCTTATCAAAAACGATTGCT 2993

Qy         64 AGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACA 123
                |   ||   ||  |  |||   ||||||  | | ||  ||||||| || |||
Db       2994 TCAGTTGAAAAGCAAACTAATAAAGAATTTGAGTATCTAATTATTGATGGTGGTTCTGAG 3053

Qy        124 GATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAA 183
              |||    | |     |    |  ||     | |   ||    |  | |    | | ||||
Db       3054 GATAATTCAA---TAGAACATGCTAAGAAAAGTGTAATATGCACAAAAATAATATCTGAA 3110

Qy        184 AAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTA 243
              |||||| | || || |||||||| |||||||| ||         | || ||||| || |
Db       3111 AAAGATAATGGAATCTATGATGCAATGAATAAAGGAGCCTCACTTGCTAAAGGTGAGTAT 3170

Qy        244 ATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCA 298
              ||   || |  |||| | || |||    |||    | |||    | ||| | | |
Db       3171 ATCATTTTTACAAACTCTGGTGATGAATTAAGAAAAAATAGCATTGAAATTTTGA 3225
```

RESULT 7
ACA54039
ID ACA54039 standard; DNA; 744 BP.
XX
AC ACA54039;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #35696.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Yersinia pestis.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU50169.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 41909; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 744 BP; 247 A; 96 C; 144 G; 257 T; 0 U; 0 Other;

Query Match 9.8%; Score 73.4; DB 8; Length 744;
Best Local Similarity 54.7%; Pred. No. 1.6e-05;
Matches 146; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

```
Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||| |||    |||| ||||  || |||||||  ||    |    |      | ||  ||
Db          1 ATGAAGATAAGTATTATTACTGCTACATATAATAGTGAGAGAACAATTTCGGATACCCTA 60
```

```
Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
Db         61 TCTTCTTTGGAGAAACAAACCTATCTGGATGTTGAGTATATTATTGTTGATGGTGCATCA 120

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
Db        121 AAGGATAATACATTAACTGTTATTAGTCAGAATTCTACGCGAGTCACTAAGATTATTAGT 180

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
Db        181 GAGTCTGATCAAGGTATTTATGATGCTTTGAATAAAGGGATAGATTTGGCAACAGGTGAC 240

Qy        241 TTAATACATTATTTAAACGCCGGCGAT 267
Db        241 ATTATTGGTTTTTTACATTCTGATGAT 267


RESULT 8
ACL37108/c
ID   ACL37108 standard; cDNA; 2000 BP.
XX
AC   ACL37108;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Rice stress-regulated promoter SEQ ID NO:15671.
XX
KW   ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW   agriculture.
XX
OS   Oryza sativa.
XX
PN   WO2003008540-A2.
XX
PD   30-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US019668.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   24-AUG-2001; 2001US-0314662P.
PR   26-SEP-2001; 2001US-0325277P.
PR   21-NOV-2001; 2001US-0332132P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Kreps J,  Briggs SP,  Cooper B,  Glazebrook J,  Goff SA,  Katagiri F;
PI   Moughamer T,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-248011/24.
XX
PT   New stress-responsive nucleic acid, useful for altering the
PT   responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT   stress, salt stress or osmotic stress.
XX
PS   Claim 48; SEQ ID NO 15671; 89pp; English.
XX
CC   The invention relates to novel abiotic stress responsive polynucleotides
CC   and polypeptides. Also disclosed are vectors, expression cassettes, host
CC   cells, and plants containing such polynucleotides. Also disclosed are
CC   methods for using the polynucleotides and polypeptides to alter the
CC   responsiveness of a plant to abiotic stress. The invention is useful in
CC   agriculture. The nucleic acid is useful for determining whether a test
CC   plant has been exposed to an abiotic stress condition. It is also useful
CC   for selecting an agent that alters abiotic stress regulated
CC   polynucleotide expression in a plant cell, and to identify a homolog or
CC   ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC   molecule and the polypeptide encoded by it are useful in altering the
CC   responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC   stress, osmotic stress or any of their combinations. The present sequence
CC   is used in the exemplification of the invention
XX
SQ   Sequence 2000 BP; 510 A; 350 C; 268 G; 518 T; 0 U; 354 Other;

  Query Match             8.5%;  Score 63.6;  DB 11;  Length 2000;
  Best Local Similarity   12.1%;  Pred. No. 0.0017;
  Matches   56;  Conservative  228;  Mismatches  177;  Indels    3;  Gaps    2;

Qy        240 CTTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAA 299
Db        486 CTGARACCAWRTKWCKWMMRKTGKKRWWYKGMKRMARKAWGKARWAYCKRTTAYATMTTT 427

Qy        300 AGAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTC 359
Db        426 TRRTRAAWWMYTGMRKRRRSMTRWRMTSRWTTWSWYKWYYYWTRRYRASRARTKGKATKY 367

Qy        360 TATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTC 419
Db        366 TMTWKWMMGCWSYTKARWWMMAYKKSWKMAWYAW-TKGTRRYYYRMKRAAYWTRWMATRS 308

Qy        420 AGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCC 479
Db        307 ATWCTGWTYWTSAKTTKWKRWAGYAKATTMKMRKTYYAGGRWWWYKKWASSRTWKWCWYK 248

Qy        480 TGAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGG--GG 537
Db        247 MYMWATKKSWWSWATTWAWKSMWYKKKWTYYTYMTSTYWTTWMWMRCRAGWWWKGGMAKR 188

Qy        538 GGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAA 597
Db        187 AAARTKWMRTYRMKKMWRYRWYKWSWARGWRAWRKRYYTYWWYMMYRCAWMWYSAYWRSW 128

Qy        598 AAAAATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCT 657
Db        127 RRWTTTKYSKMYYWKMWMMRRTYARAKAYSRWWMWKKRCRTRWKGWTRWKTWKMMWRWYK 68

Qy        658 GAACGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCT 701
Db         67 KRTYRMKKSAKWWYMRSGAYRRRWMRRMAYSRRRSWMRKTSYYY 24


RESULT 9
AEA29447
ID   AEA29447 standard; DNA; 13749 BP.
XX
AC   AEA29447;
XX
DT   28-JUL-2005  (first entry)
XX
DE   Escherichia coli 0159 O-antigen DNA, SEQ ID 1.
XX
KW   antigen; Escherichia coli 0159; ds.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1021. .2400
FT                   /*tag=  a
FT                   /product= "Escherichia coli 0159 O-antigen orf1"
FT   CDS             2397. .3134
FT                   /*tag=  b
FT                   /product= "Escherichia coli 0159 O-antigen orf2"
FT   CDS             3106. .3942
FT                   /*tag=  c
FT                   /product= "Escherichia coli 0159 O-antigen orf3"
FT   CDS             3929. .4942
FT                   /*tag=  d
FT                   /product= "Escherichia coli 0159 O-antigen orf4"
FT   CDS             4935. .6080
FT                   /*tag=  e
FT                   /product= "Escherichia coli 0159 O-antigen orf5"
FT   CDS             6077. .7195
FT                   /*tag=  f
FT                   /product= "Escherichia coli 0159 O-antigen orf6"
FT   CDS             7199. .8164
FT                   /*tag=  g
FT                   /product= "Escherichia coli 0159 O-antigen orf7"
```

```
FT   CDS             8167. .8628
FT                   /*tag=  h
FT                   /product= "Escherichia coli 0159 O-antigen orf8"
FT   CDS             8634. .10037
FT                   /*tag=  i
FT                   /product= "Escherichia coli 0159 O-antigen orf9"
FT   CDS             10037. .10783
FT                   /*tag=  j
FT                   /product= "Escherichia coli 0159 O-antigen orf10"
FT   CDS             10789. .12213
FT                   /*tag=  k
FT                   /product= "Escherichia coli 0159 O-antigen orf11"
XX
PN   CN1546511-A.
XX
PD   17-NOV-2004.
XX
PF   03-DEC-2003; 2003CN-01107161.
XX
PR   03-DEC-2003; 2003CN-01107161.
XX
PA   (UYNA-) UNIV NANKAI.
XX
PI   Feng L,  Kong Q,  Guo H;
XX
DR   WPI; 2005-164293/18.
XX
PT   Nucleotide specific for Escherichia coli 0159 O-antigen.
XX
PS   Claim 1; SEQ ID NO 1; 29pp; Chinese.
XX
CC   The invention relates to a novel nucleotide specific for an Escherichia
CC   coli 0159 O-antigen, which is the total nucleotide sequence of the gene
CC   cluster for controlling the synthesis of O-antigens in Escherichia coli
CC   0159. For example, the isolated nucleotide represented by SEQ ID No:1,
CC   with overall length of 13749 bases, or nucleotide of SEQ ID No:1
CC   including one or more inserted, deleted or substituted bases and sustains
CC   the isolated nucleotide functions, it also includes the oligonucleotides
CC   of glycosyl transferase genes and oligosaccharide unit treatment genes in
CC   the O-antigen gene cluster originating from Escherichia coli 0159. The
CC   invention proves the high specificity of oligonucleotides for Escherichia
CC   coli 0159 O-antigen through PCR. A method for detecting and identifying
CC   Escherichia coli 0159 in the human body and in the environment by means
CC   of the oligonucleotide is also disclosed in the invention. This
CC   polynucleotide sequence represents the DNA of an Escherichia coli 0159 O-
CC   antigen of the invention.
XX
SQ   Sequence 13749 BP; 4189 A; 2306 C; 2814 G; 4440 T; 0 U; 0 Other;

  Query Match             8.4%;  Score 62.8;  DB 14;  Length 13749;
  Best Local Similarity   51.4%;  Pred. No. 0.003;
  Matches  145;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||| ||| ||  |||| || |  |||||||| || ||   ||      |  || |||
Db      10037 ATGCGAATTTCTATTATCACAGCTACTTATAGTAGTGAAAAAACTCTCTTTGATACATTA 10096

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
                | | ||  ||     ||         | | ||||   |||  | || ||||| |  ||
Db      10097 CTTTCTCTAGAAAAGCAAACGCATCCAGATATTGAATATATAGTTGTAGATGGAGCATCA 10156

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              | ||||   || | |    | ||||  ||  || ||  |    |  | ||   |||
Db      10157 AAAGATAATACTATTAAGCTGATTAAGAGTAATAGTACAAAAGTTTCAAAAATCATTTGC 10216

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              |||  ||| | || || |||||||  |||||||  ||          ||   || ||
Db      10217 GAACCCGATAAGGGCATTTATGATGCTCTGAATAAAGGTATTCAAGCAGCTTCGGGTGAT 10276

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGAT 282
              | ||   || |||| |  | |  |||    ||  |  |||
Db      10277 GTTATTGGTTTTTTACATTCTGATGATTTACTAGCTTATGAT 10318
```

```
RESULT 10
ABX66419/c
ID   ABX66419 standard; DNA; 1024 BP.
XX
AC   ABX66419;
XX
DT   07-MAY-2003  (first entry)
XX
DE   Helicobacter pylori selected interacting domain (SID) DNA #1018.
XX
KW   Protein-protein interaction; ulcer; selected interacting domain; SID;
KW   gene; ds.
XX
OS   Helicobacter pylori.
XX
PN   WO200266501-A2.
XX
PD   29-AUG-2002.
XX
PF   28-DEC-2001; 2001WO-EP015428.
XX
PR   02-JAN-2001; 2001US-0259302P.
XX
PA   (HYBR-) HYBRIGENICS.
PA   (INSP ) INST PASTEUR.
XX
PI   Legrain P,  Rain J,  Colland F,  De Reuse H,  Labigne A;
XX
DR   WPI; 2002-674910/72.
DR   P-PSDB; ABU51675.
XX
PT   New complexes of protein-protein interactions in Helicobacter pylori,
PT   useful for identifying modulating compounds for treating or preventing
PT   ulcers in mammals.
XX
PS   Claim 7; Page 335; 642pp; English.
XX
CC   The invention describes a complex of protein-protein interactions in
CC   Helicobacter pylori selected from 421 complexes given in the
CC   specification. The complex of protein-protein interactions are useful for
CC   screening for agents which modulate the interaction of proteins.
CC   Modulating compounds which binds to a targeted bacterial protein may be
CC   used for treating or preventing ulcers in a human or animal. This
CC   sequence encodes a selected interacting domain (SID), identified via
CC   protein-protein interactions
XX
SQ   Sequence 1024 BP; 292 A; 235 C; 148 G; 349 T; 0 U; 0 Other;

  Query Match             7.6%;  Score 56.4;  DB 6;  Length 1024;
  Best Local Similarity   50.8%;  Pred. No. 0.045;
  Matches  135;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy          2 TGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAA 61
              | ||| | || || || || | |  || |||||  |   ||    | ||| | ||  |
Db        943 TAAAAGTTTCTGTGATCACGGCGTGTTTTAATAGCGAAAAAACCATTGAAGACACCATTC 884

Qy         62 GTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTA 121
              |     |     | ||     |||    ||||    || ||  | || || |      |
Db        883 TTTCCGTGCTTCATCAAACTTATAAAAACATTGAATACATCATTATAGACGGGGCTAGCA 824

Qy        122 CAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATG 181
              | ||| | ||         |||||   | |  ||    ||  |     |      |  ||
Db        823 CGGATAGCACTTTAGAAATCATTCAAAAACATAGAGATAAAATCGCTTGCGTAATGAGTG 764

Qy        182 AAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACT 241
              |||||||||| || || ||||| || |||||||||||||  ||  |     | |  || ||
Db        763 AAAAAGATGAGGGCATTTATGACGCTATGAATAAGGGCATAAAGCGTTCTAGTGGGGATA 704

Qy        242 TAATACATTATTTAAACGCCGGCGAT 267
              | ||   ||  || ||   ||  |||
Db        703 TTATCGCTTTATTGAATAGCGATGAT 678
```

```
RESULT 11
AEA39955
ID   AEA39955 standard; DNA; 9415 BP.
XX
AC   AEA39955;
XX
DT   11-AUG-2005  (first entry)
XX
DE   Escherichia coli O-antigen biosynthesis gene.
XX
KW   O-antigen; biosynthesis; glycosyltransferase; gene; ds.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1142. .2389
FT                   /*tag=  a
FT                   /note= "ORF1 (wzx). Specifically claimed in Claim 3"
FT   CDS             2401. .3498
FT                   /*tag=  b
FT                   /note= "ORF2 (glf). Specifically claimed in Claim 3"
FT   CDS             3495. .4409
FT                   /*tag=  c
FT                   /note= "ORF3. Specifically claimed in Claim 3"
FT   CDS             4447. .5487
FT                   /*tag=  d
FT                   /note= "ORF4 (wzy). Specifically claimed in Claim 3"
FT   CDS             5703. .6824
FT                   /*tag=  e
FT                   /note= "ORF5. Specifically claimed in Claim 3"
FT   CDS             6802. .7764
FT                   /*tag=  f
FT                   /note= "ORF6. Specifically claimed in Claim 3"
FT   CDS             7779. .8702
FT                   /*tag=  g
FT                   /note= "ORF7. Specifically claimed in Claim 3"
XX
PN   CN1539847-A.
XX
PD   27-OCT-2004.
XX
PF   24-OCT-2003; 2003CN-01106708.
XX
PR   24-OCT-2003; 2003CN-01106708.
XX
PA   (UYNA-) UNIV NANKAI.
XX
PI   Feng L,  Wang W,  Xia Q;
XX
DR   WPI; 2005-123678/14.
XX
PT   Nucleotide specific to 0 antigen of coliform 096 type.
XX
PS   Claim 1; SEQ ID NO 1; 33pp; Chinese.
XX
CC   The invention relates to a nucleotide specific to the O-antigen of
CC   Escherichia coli 096. The invention discloses a complete nucleotide
CC   sequence (SEQ ID No:1) of the gene for controlling the synthesis of the
CC   said O-antigen in E. coli 096. It also includes the oligonucleotide
CC   coming from the glycosyltransferase gene and oligose unit processing gene
CC   of E. coli 096. The process for separating the said complete sequence and
CC   the method for using the said oligonucleotide to detect Escherichia coli
CC   96 are also disclosed. This sequence represents the E. coli O-antigen
CC   biosynthesis gene of the invention.
XX
SQ   Sequence 9415 BP; 3018 A; 1390 C; 1827 G; 3180 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 55.8;  DB 14;  Length 9415;
  Best Local Similarity   54.0%;  Pred. No. 0.075;
  Matches  114;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy         86 AACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACAGATGGAACGAATCGTGTCATTA 145
              || | | || ||||| ||  |||||  || || || |||||    | |  ||  |||
Db       6899 AAGATATAGAAATTATCATTATTGATGATGGTTCAACGGATGGTTTAAGTAATGCTATTT 6958

Qy        146 GTAGATTTACTAGTATGAATATTACACATGTTTATGAAAAAGATGAAGGGATATATGATG 205
               |  || ||   | |  || || | | || || |  || || |||  ||  | |    ||
Db       6959 CTTCATATAAAGGAAACAAAATAATATATTTTCACAAAGAAAATGGTGGTGTTTCATCTG 7018

Qy        206 CGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTAATACATTATTTAAACGCCGGCG 265
              | |  ||||||||       || |||| ||| | |   | | || |||  | |  | |
Db       7019 CAAGAAATAAGGGAATTGATTTAGCCACAGGTAAATATTTGCTTTTTTTTGATGGAGATG 7078

Qy        266 ATAGCGTAATTGGAGATATATATAAAAATAT 296
              ||  |      |  ||| |   |||  || |
Db       7079 ATTACTATTCCGAGGATCTCATTAATTATCT 7109


RESULT 12
AEA33353
ID   AEA33353 standard; DNA; 15390 BP.
XX
AC   AEA33353;
XX
DT   28-JUL-2005  (first entry)
XX
DE   E. coli O156 O-antigen sequence, seq id 1.
XX
KW   Microorganism detection; O-antigen; gene; ds.
XX
OS   Escherichia coli; O156.
XX
PN   CN1546512-A.
XX
PD   17-NOV-2004.
XX
PF   03-DEC-2003; 2003CN-01107162.
XX
PR   03-DEC-2003; 2003CN-01107162.
XX
PA   (UYNA-) UNIV NANKAI.
XX
PI   Feng L,  Kong Q,  Guo H;
XX
DR   WPI; 2005-164294/18.
XX
PT   Polynucleotide specific for Escherichia coli 0156 O-antigen.
XX
PS   Claim 1; SEQ ID NO 1; 31pp; Chinese.
XX
CC   The invention provides a polynucleotide specific for the Escherichia coli
CC   0156 O-antigen, which is the total nucleotide sequence of the gene
CC   cluster for controlling the synthesis of O-antigens in E. coli 0156. It
CC   includes the glycosyl transferase genes and oligosaccharide unit
CC   treatment genes in the O-antigen gene cluster originated from E. coli
CC   0156. The invention includes detecting and identifying E. coli 0156 in
CC   the human body and in the environment (e.g. by PCR). The current sequence
CC   represents the E. coli O156 O-antigen sequence.
XX
SQ   Sequence 15390 BP; 4762 A; 2581 C; 3133 G; 4914 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 55.8;  DB 14;  Length 15390;
  Best Local Similarity   50.6%;  Pred. No. 0.079;
  Matches  135;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              |||  ||| ||  |||| || |  |||||||||| ||   ||     |      | || |||
Db      11707 ATGCGAATTTCTATTATCACAGCTACTTATAATAGTGAAAAAACTCTCCTTGATACATTA 11766

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
                |  | ||  ||     ||       |  | | ||   |||  | || ||||| |  ||
Db      11767 CTTTCTCTAGAAAAACAAACGTATCCAGATATCGAATATATAGTTGTAGATGGAGCATCG 11826

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              |  |||   || | |    | || |   |   |  ||  |    |   | |    |||
```

```
Db      11827 AACGATAATACAATTAAGCTGATCAAGGGTAATTGTACAAAAGTTTCAAAAATCATTTGC 11886

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240

Db      11887 GAACCCGATAAGGGCATTTATGATGCTCTGAATAAAGGTATTCAAGCAGCTTCGGGTGAT 11946

Qy        241 TTAATACATTATTTAAACGCCGGCGAT 267

Db      11947 GTTATTGGTTTTTTACATTCTGATGAT 11973
```

RESULT 13
AAX06748
ID AAX06748 standard; DNA; 14516 BP.
XX
AC AAX06748;
XX
DT 26-APR-1999 (first entry)
XX
DE E. coli O111 antigen gene cluster.
XX
KW O antigen; O111 antigen; diagnosis; wbdH gene; wzx gene; wzy gene;
KW wbdM gene; rfb gene; transferase; polymerase; flippase; diarrhoea;
KW haemorrhagic colitis; ds.
XX
OS Escherichia coli.
XX

FH	Key	Location/Qualifiers
FT	CDS	739. .1932
FT		/*tag= a
FT		/product= "galactosyl transferase"
FT		/note= "wbdH gene (ORF1), this region is specifically
FT		claimed in Claim 6, and encodes AAW88300"
FT	CDS	1983. .3101
FT		/*tag= b
FT		/note= "gmd gene (ORF2); encodes AAW88301"
FT	CDS	3174. .3623
FT		/*tag= c
FT		/note= "wbdI gene (ORF3); encodes AAW88302"
FT	CDS	3657. .5105
FT		/*tag= d
FT		/note= "manC gene (ORF4); encodes AAW88303"
FT	CDS	5114. .6484
FT		/*tag= e
FT		/note= "manB gene (ORF5); encodes AAW88304"
FT	CDS	6521. .7444
FT		/*tag= f
FT		/note= "wbdJ gene (ORF6); encodes AAW88305"
FT	CDS	7441. .8607
FT		/*tag= g
FT		/note= "wbdK gene (ORF7); encodes AAW88306"
FT	CDS	8646. .9911
FT		/*tag= h
FT		/product= "O antigen flippase"
FT		/note= "wzx gene (ORF8), this region is specifically
FT		claimed in Claim 6; encodes AAW88307"
FT	CDS	9901. .10953
FT		/*tag= i
FT		/product= "O antigen polymerase"
FT		/note= "wzy gene (ORF9), this region is specifically
FT		claimed in Claim 6; encodes AAW88308"
FT	CDS	10931. .11824
FT		/*tag= j
FT		/product= "colitose or glucose transferase"
FT		/note= "wbdL gene (ORF10); encodes AAW88309"
FT	CDS	11821. .12945
FT		/*tag= k
FT		/product= "colitose or glucose transferase"
FT		/note= "wbdM gene (ORF11), this region is specifically
FT		claimed in Claim 6; encodes AAW88310"
FT	CDS	13124. .14516
FT		/*tag= l
FT		/note= "gnd gene (partial sequence); encodes AAW88311"

XX
PN WO9850531-A1.
XX
PD 12-NOV-1998.
XX
PF 01-MAY-1998; 98WO-AU000315.
XX
PR 01-MAY-1997; 97AU-00006545.
PR 22-JUL-1997; 97AU-00008162.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Reeves PR, Wang L;
XX
DR WPI; 1999-059669/05.
DR P-PSDB; AAW88300, AAW88301, AAW88302, AAW88303, AAW88304, AAW88305,
DR AAW88306, AAW88307, AAW88308, AAW88309, AAW88310, AAW88311.
XX
PT Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT useful for detecting specific strains in, e.g. food, faeces or patient
PT samples.
XX
PS Claim 6; Fig 7; 165pp; English.
XX
CC This is the nucleotide sequence of a gene cluster involved in the
CC biosynthesis of the Escherichia coli O111 O antigen. The gene cluster was
CC isolated from a cosmid gene bank of E. coli M92 chromosomal DNA by
CC screening with E. coli O111 antiserum. Open reading frames within the
CC gene cluster were identified on the basis of homology to known sequences.
CC The genes encode proteins (see AAW88300-11) involved the synthesis of
CC sugars present in the polysaccharide antigen, and in the transport or
CC processing of polysaccharide or oligosaccharide units. The use of nucleic
CC acid molecules derived from particular assembly and transport genes,
CC particularly wbd (transferase), wzx (flippase) and wzy (polymerase)
CC genes, within O antigen gene clusters improves the specificity of methods
CC for the detection and identification of O antigens, e.g. in testing food-
CC or faecal-derived samples, or samples from patients. The O antigen is a
CC major virulence factor of enteropathogenic E. coli strains that cause
CC diarrhoea and haemorrhagic colitis
XX
SQ Sequence 14516 BP; 4582 A; 2221 C; 2971 G; 4742 T; 0 U; 0 Other;

Query Match 7.4%; Score 55.2; DB 2; Length 14516;
Best Local Similarity 50.8%; Pred. No. 0.1;
Matches 132; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```
Qy          8 TATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGTAGTT 67

Db      10989 TTTCAATAATCATTGCAACTTATAATTCTGAACTTGATATAGCTAAGTGTTTGCAATCGG 11048

Qy         68 TATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACAGATG 127

Db      11049 TAACTAATCAATCTTATAAGAATATTGAAATCATAATAATGGATGGAGGATCTTCTGATA 11108

Qy        128 GAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAAAAAG 187

Db      11109 AAACGCTTGATATTGCAAAATCGTTTAAAGACGACCGAATAAAAATAGTTTCAGAGAAAG 11168

Qy        188 ATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTAATAC 247

Db      11169 ATCGTGGAATTTATGATGCCTGGAATAAAGCAGTTGATTTATCCATTGGTGATTGGGTAG 11228

Qy        248 ATTATTTAAACGCCGGCGAT 267

Db      11229 CATTTATTGGTTCAGATGAT 11248
```

RESULT 14
AAZ56375
ID AAZ56375 standard; DNA; 14516 BP.
XX
AC AAZ56375;
XX

DT 17-MAR-2000 (first entry)
XX
DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:45.
XX
KW Flagellin; fliC; antigen; detection; ds.
XX
OS Escherichia coli.
XX
PN WO9961458-A1.
XX
PD 02-DEC-1999.
XX
PF 21-MAY-1999; 99WO-AU000385.
XX
PR 21-MAY-1998; 98AU-00003634.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Reeves PR, Wang L;
XX
DR WPI; 2000-072598/06.
XX
PT Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, feces, etc.
XX
PS Claim 3; Page 216-221; 245pp; English.
XX
CC AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all or
CC part of an Escherichia coli flagellin protein except a protein expressed
CC by E. coli H1, H7, H12 or H48 type strains. The present invention also
CC describes a method of detecting the presence of E. coli of a particular H
CC serotype in a sample, comprising specifically hybridising a nucleic acid,
CC preferably at least a pair, derived from a flagellating gene, specific
CC for a particular flagellin gene associated with the H serotype, to any
CC E.coli in the sample which contain the gene, and detecting any hybridised
CC molecules, identifying the presence of that serotype in the sample. (I)
CC are useful for: (1) detecting the presence of E. coli of H serotype in a
CC sample by hybridising at least one or a pair of (I) to any E. coli in the
CC sample and detecting the hybridised nucleic acid molecules; and (2) for
CC detecting the presence of both O and H-serotypes of E. coli by
CC hybridising at least one or a pair of (I) to any E. coli present in the
CC sample and detecting the hybridised nucleic acid molecules. (I) is
CC particularly useful for detecting the combination of O and H antigen.
CC Hybridised (I) when using at least one (I) is detected by southern blot
CC analysis and, when using a pair of (I), is detected by polymerase chain
CC reaction (PCR). AAZ56399 to AAZ56420 represent primers used in the
CC exemplification of the present invention
XX
SQ Sequence 14516 BP; 4582 A; 2221 C; 2971 G; 4742 T; 0 U; 0 Other;

Query Match 7.4%; Score 55.2; DB 3; Length 14516;
Best Local Similarity 50.8%; Pred. No. 0.1;
Matches 132; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

Qy 8 TATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGTAGTT 67
Db 10989 TTTCAATAATCATTGCAACTTATAATTCTGAACTTGATATAGCTAAGTGTTTGCAATCGG 11048

Qy 68 TATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACAGATG 127
Db 11049 TAACTAATCAATCTTATAAGAATATTGAAATCATAATAATGGATGGAGGATCTTCTGATA 11108

Qy 128 GAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAAAAAG 187
Db 11109 AAACGCTTGATATTGCAAAATCGTTTAAAGACGACCGAATAAAAATAGTTTCAGAGAAAG 11168

Qy 188 ATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTAATAC 247
Db 11169 ATCGTGGAATTTATGATGCCTGGAATAAAGCAGTTGATTTATCCATTGGTGATTGGGTAG 11228

Qy 248 ATTATTTAAACGCCGGCGAT 267
Db 11229 CATTTATTGGTTCAGATGAT 11248

RESULT 15
AEA31080
ID AEA31080 standard; DNA; 10432 BP.
XX
AC AEA31080;
XX
DT 28-JUL-2005 (first entry)
XX
DE Shigella boydii 17 O-antigen gene cluster DNA.
XX
KW microorganism detection; O-antigen; operon; ds.
XX
OS Shigella boydii.
XX
PN CN1563065-A.
XX
PD 12-JAN-2005.
XX
PF 19-APR-2004; 2004CN-00019050.
XX
PR 19-APR-2004; 2004CN-00019050.
XX
PA (TIAN-) TIANJIN BIOCHIP TECH CO LTD.
XX
PI Wang L, Yang J, Feng L;
XX
DR WPI; 2005-307140/32.
XX
PT Nucleotide peculiar to O-antigen of Shigella boydii 17 useful for
PT controlling synthesis of O-antigen and detecting and identifying Shigella
PT in human body.
XX
PS Claim 1; SEQ ID NO 1; 35pp; Chinese.
XX
CC The invention relates to a novel Shigella boydii 17 O-antigen
CC polynucleotide comprising the full-length nucleotide sequence of the gene
CC cluster responsible for controlling O-antigen synthesis in Shigella
CC boydii 17. The cluster includes the glycosyl transferase genes and
CC oligosaccharide unit treatment genes. The polynucleotide of the invention
CC may be used for detecting the O-antigen and thus the Shigella boydii 17
CC bacterium in the human body or in the environment, using methods such as
CC PCR. The current sequence is that of the Shigella boydii 17 O-antigen
CC gene cluster DNA of the invention which contains the galF, wzx, orf2,
CC orf3, wzy, orf5, gne, orf7 and gnd genes.
XX
SQ Sequence 10432 BP; 3366 A; 1592 C; 2059 G; 3415 T; 0 U; 0 Other;

Query Match 7.1%; Score 53.2; DB 14; Length 10432;
Best Local Similarity 50.0%; Pred. No. 0.26;
Matches 133; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy 1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
Db 5912 ATGTTTTTAACAGTAGTTATTCCGCTTTATAATAAAAGTCAATCAATTGAGTCCACGATT 5971

Qy 61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
Db 5972 AAATCTGTTCTAAATCAAACGCATACTGATTTCGAATTAATTGTAGTAAATGATGGTTCA 6031

Qy 121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
Db 6032 ACGGACGGAAGTGAAAATGTAATAAAACAGTTTACTGATGAACGTATAAAACTTTTGAAT 6091

Qy 181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
Db 6092 ATTACAAACTCTGGAGTATCAGTTGCTAGAAATACCGGTATAAGATCAGCGAAATCAAGC 6151

Qy 241 TTAATACATTATTTAAACGCCGGCGA 266
Db 6152 TATGTGTGTTTATTAGATGCAGATGA 6177

Search completed: May 10, 2006, 10:59:15
Job time : 351.71 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 11:00:02 ; Search time 2392.49 Seconds
(without alignments)
14608.170 Million cell updates/sec

Title: US-09-423-093B-2_COPY_2011_2757
Perfect score: 747
Sequence: 1 atgaaaatatcagttataac.....caccatatgataatgaataa 747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

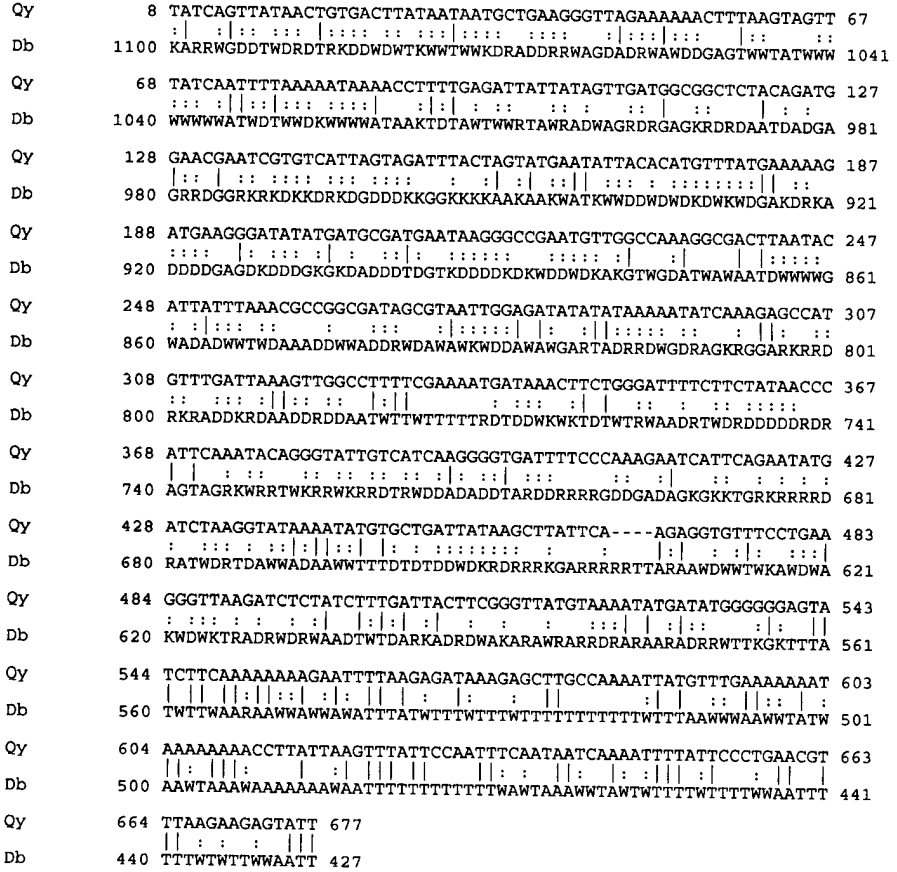
SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	76.4	10.2	1101	10	CNS0039G	AL063921 Drosophil
	2	65.4	8.8	1101	10	CNS00EVL	AL069706 Drosophil
	3	64.2	8.6	1100	1	AJ926153	AJ926153 AJ926153
c	4	61.6	8.2	1101	10	CNS00EVL	AL069706 Drosophil
c	5	60.6	8.1	1200	10	CNS016CO	AL106578 Drosophil
	6	60	8.0	1092	10	CNS020K7	AL175696 Tetraodon
c	7	60	8.0	1101	10	CNS017KX	AL108171 Drosophil
	8	58	7.8	1101	10	CNS017ZQ	AL108704 Drosophil
c	9	57.8	7.7	851	10	DU074282	DU074282 69279 Tom
c	10	57.4	7.7	1101	10	CNS0039Q	AL063931 Drosophil
	11	57	7.6	1101	10	CNS00YZS	AL097042 Drosophil
	12	56.6	7.6	1200	10	CNS016CO	AL106578 Drosophil
	13	56.6	7.6	1367	8	DN683496	DN683496 CGX30-C12
c	14	55.8	7.5	1047	10	CW953272	CW953272 TcB39.2_C
c	15	55	7.4	878	7	CO389521	CO389521 AGENCOURT
	16	55	7.4	1168	10	CL078758	CL078758 CH216-153
	17	54.2	7.3	500	1	AU088558	AU088558 AU088558
c	18	54	7.2	822	10	CNS009CW	AL053618 Drosophil
	19	53.8	7.2	905	10	CNS00KHX	AL077798 Drosophil
	20	53.8	7.2	1011	8	DN581585	DN581585 92334661
	21	53.8	7.2	1077	1	AJ927831	AJ927831 AJ927831
	22	53.8	7.2	1101	10	CNS00FMC	AL070972 Drosophil
c	23	53.8	7.2	1204	10	CNS016E2	AL106628 Drosophil
c	24	53.4	7.1	1201	10	CNS0167M	AL106396 Drosophil
c	25	53.2	7.1	1101	10	CNS001FB	AL060732 Drosophil
	26	53.2	7.1	1101	10	CNS00EO7	AL069440 Drosophil
	27	53	7.1	499	1	AU087437	AU087437 AU087437
c	28	52.8	7.1	786	10	CNS009A5	AL053317 Drosophil
c	29	52.8	7.1	1094	10	CNS012FZ	AL101513 Drosophil
	30	52.8	7.1	1136	1	AJ926338	AJ926338 AJ926338
	31	52.6	7.0	939	5	BY720774	BY720774 BY720774
c	32	52.6	7.0	1101	10	CNS002DB	AL062360 Drosophil
c	33	52.6	7.0	1274	7	CR553336	CR553336 CR553336
c	34	52.4	7.0	975	9	BH179465	BH179465 014_P_10-
c	35	52.2	7.0	932	6	CD390089	CD390089 AGENCOURT
c	36	52.2	7.0	987	10	CNS014PQ	AL104456 Drosophil
	37	52.2	7.0	1101	10	CNS0182P	AL108811 Drosophil
	38	52.2	7.0	1132	9	BZ695089	BZ695089 SP__Ba005
	39	52	7.0	976	10	CL516956	CL516956 SAIL_98_C
	40	52	7.0	1101	10	CNS0039G	AL063921 Drosophil
	41	52	7.0	1101	10	CNS00EQL	AL069526 Drosophil
	42	52	7.0	1146	10	CNS021G2	AL176843 Tetraodon
c	43	52	7.0	1242	10	CL068807	CL068807 CH216-115
c	44	51.8	6.9	867	11	CNS075BG	AL429890 clone BA0
	45	51.8	6.9	895	10	CZ536560	CZ536560 SRAA-aad0

ALIGNMENTS

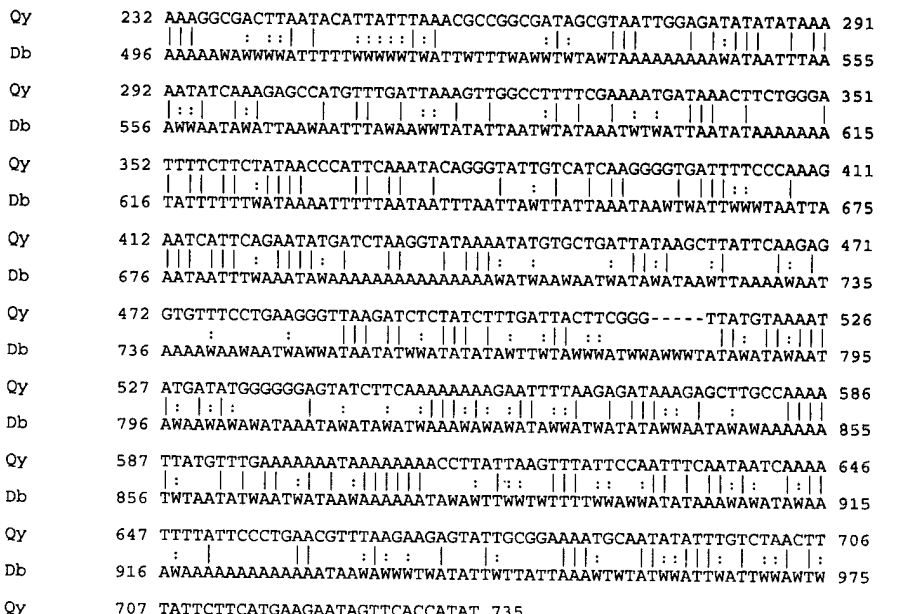
RESULT 1
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

Query Match 10.2%; Score 76.4; DB 10; Length 1101;
Best Local Similarity 17.7%; Pred. No. 3.5e-05;
Matches 119; Conservative 288; Mismatches 263; Indels 4; Gaps 1;



RESULT 2
CNS00EVL
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"
ORIGIN

Query Match 8.8%; Score 65.4; DB 10; Length 1101;
Best Local Similarity 34.8%; Pred. No. 0.0057;
Matches 177; Conservative 83; Mismatches 244; Indels 5; Gaps 1;



RESULT 3
AJ926153
LOCUS AJ926153 1100 bp mRNA linear EST 10-JUN-2005
DEFINITION AJ926153 Theileria annulata merozoite Theileria annulata cDNA clone tam013a07_q1k, mRNA sequence.
ACCESSION AJ926153
VERSION AJ926153.1 GI:67496527

```
KEYWORDS    EST.
SOURCE      Theileria annulata
  ORGANISM  Theileria annulata
            Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
            Theileria.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Pain,A., Renauld,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W.,
            Kerhornou,A., Aslett,M., Bishop,R., Bouchier,C., Cochet,M.,
            Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Fosker,N.,
            Gardner,M., Goble,A., Griffiths-Jones,S., Harris,D.E., Katzer,F.,
            Larke,N., Lord,A., Maser,P., McKellar,S., Mooney,P., Morton,F.,
            Nene,V., O'Neil,S., Price,C., Quail,M.A., Rabbinowitsch,E.,
            Rawlings,N.D., Rutter,S., Saunders,D., Seeger,K., Shah,T.,
            Squares,R., Squares,S., Tivey,A., Walker,A.R., Woodward,J.,
            Dobbelaere,D.A.E., Langsley,G., Rajandream,M.-A., McKeever,D.,
            Shiels,B., Tait,A., Barrell,B. and Hall,N.
  TITLE     The genome of the host-cell transforming parasite Theileria
            annulata and a comparison with T. parva
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Pain A
            The Pathogen Sequencing Unit
            The Wellcome Trust Sanger Institute
            Genome Campus, CB10 1SA, UNITED KINGDOM
            Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of
            Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="Theileria annulata"
                     /mol_type="mRNA"
                     /isolate="Ankara (clone D7)"
                     /db_xref="taxon:5874"
                     /clone="tam013a07_q1k"
                     /dev_stage="merozoite"
                     /lab_host="Bos taurus (cow)"
                     /clone_lib="Theileria annulata merozoite"
                     /note="country: Turkey:Ankara"
ORIGIN

  Query Match             8.6%;  Score 64.2;  DB 1;  Length 1100;
  Best Local Similarity   44.3%;  Pred. No. 0.01;
  Matches  290;  Conservative    0;  Mismatches  356;  Indels    8;  Gaps    3;

Qy         66 TTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACAGA 125
              | ||  | || | ||  ||| |     |  | | | ||           |  |   |
Db        428 TATAATATTTATTAATTTAATANTNNNTNTATTAAATTTATAATNNNATANTTANAATTA 487

Qy        126 TGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAAAA 185
                 |    | |    | |  | ||  |  | ||    | ||  |  | ||  | |||
Db        488 NTTTATATTTTNTTAATTAANAANATANATAAATAATTTTTTTNAAATTATTNNTANAAA 547

Qy        186 AGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTAAT 245
                |  ||   |||||| ||     | | |          ||  |     |||     |  ||
Db        548 NTAAAAATTTATATATAATTTTNTTANTTTATTTTTAAAATAANTTAAATTAAATATATA 607

Qy        246 ACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAA----AATATCAAAG 301
                || | ||          | |   |||||      ||||| |||     |||||  |
Db        608 TNTTTTTATATTTAATAATAAAATTNTAATTAATATTATATAAAAAAATTAATATTNATA 667

Qy        302 AGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCTA 361
                 ||  |  ||||    |      ||||  || ||  |||| |   |   ||| | || ||
Db        668 TAAAATAATATATTATTTTNNNTTTTTTAAAATATTTTAAANTAATTTAATTATATTATA 727

Qy        362 TAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCAG 421
                | |   |||     |  |   |  | |    ||    | || ||   | | ||| | |
Db        728 TTA--TATTTNTTTNAATAATTATTTTTATAAAANATTATNTTAATATATAATAAATTTT 785

Qy        422 AATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTG 481
                | || || ||   |||| | ||         | |||     |  || || |  | ||   |
Db        786 TAAATTATATATAATATATATTAAAATAANAATATTTATTATTTAAATATTTATTAATTT 845

Qy        482 AAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGAG 541
                ||   |||| |    |   ||  | |  ||     |     | || |  |||       |
Db        846 AAATATTAAAAANATTTAATTATANTTATTAAAAATTATATATATTTTTTATAAAATAAT 905
```

```
Qy        542 TATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAA 601
              |||  | | ||  ||| |   |||| |  |||| |   |   ||||||  | ||  ||   |
Db        906 TATTATTATAATAAATATAATTAATAATTAAATAATATTTTAAAATTTTTTTATAATNTA 965

Qy        602 ATAA--AAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGA 659
                | |  || | |    || |   |||||  |  || | || |   |  |||     | |
Db        966 TTTATTAATATATANAATAATNATTATTTTATATTTANTATTTTTATAATTAAATTATAA 1025

Qy        660 ACGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTT 713
              |   || |  | | |||     |||||  |  ||  || | ||| ||| ||  |
Db       1026 AAAATTTAATATAATATATATTAAAATATATAATTNTTTTATAAATTTTTTNAT 1079

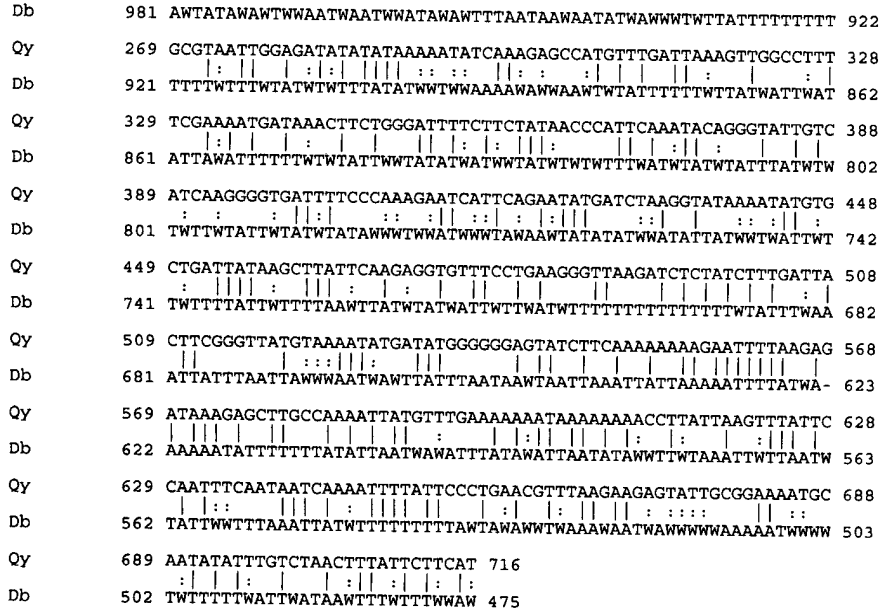

RESULT 4
CNS00EVL/c
LOCUS       CNS00EVL                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             8.2%;  Score 61.6;  DB 10;  Length 1101;
  Best Local Similarity   34.2%;  Pred. No. 0.033;
  Matches  215;  Conservative   96;  Mismatches  315;  Indels    2;  Gaps    2;

Qy         90 TTTTGAGATTATTATAGTTGATGGCGGCTCTACAGATGGAACGAATCGTGTCATTAGTAG 149
              ||:: | :|: ||:|  || :|      : || | |   ||  ::|      :| : :
Db       1101 TTWWTATWTWTTTWTTTTTWWTATWTTAWATATATAAWTAATTWWTWTATATWTAWWWTA 1042

Qy        150 ATTTACTAGTATGAATATTACACATGTTTATGAA-AAAGATGAAGGGATATATGATGCGA 208
               : ||     :|  :||  | | |: || || || | | :| :    ||||||  :   |
Db       1041 TWATATATTAWTTTWTAAAATATAWWTTAATAAATATATWTAWTAWWATATATAWATAWA 982

Qy        209 TGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTAATACATTATTTAAACGCCGGCGATA 268
                 |||:      | :  |    |:|    | | |: | ||:|::: :          |
```



RESULT 5
CNS016CO/c
LOCUS CNS016CO 1200 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN15E04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106578
VERSION AL106578.1 GI:5622626
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1200)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1200
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN15E04"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"
ORIGIN

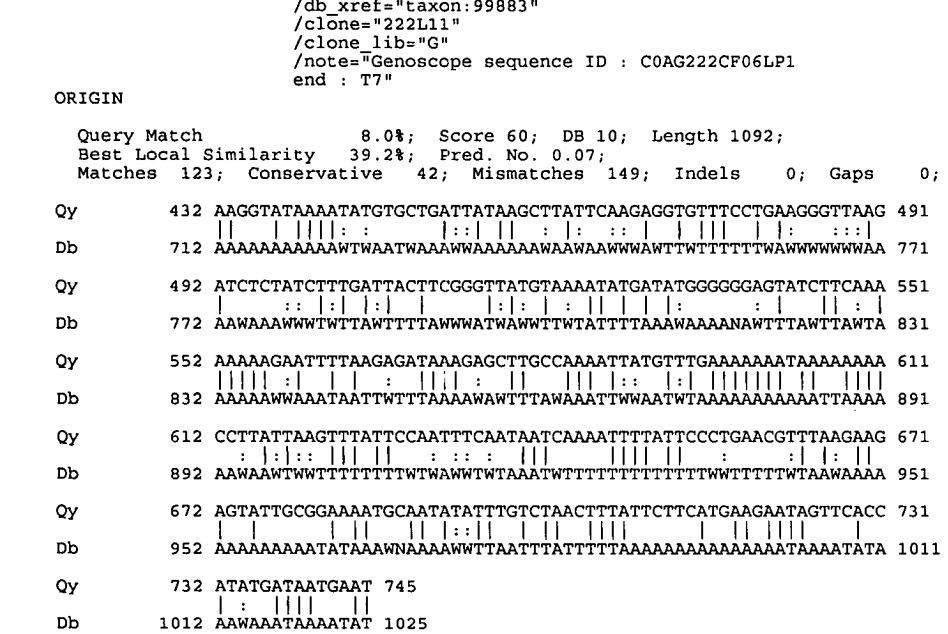
Query Match 8.1%; Score 60.6; DB 10; Length 1200;
Best Local Similarity 37.3%; Pred. No. 0.053;
Matches 116; Conservative 52; Mismatches 143; Indels 0; Gaps 0;

RESULT 6
CNS020K7
LOCUS CNS020K7 1092 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 222L11 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL175696
VERSION AL175696.1 GI:7813753
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
PUBMED 10899143
REFERENCE 3 (bases 1 to 1092)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .1092

/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="222L11"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG222CF06LP1
end : T7"

ORIGIN

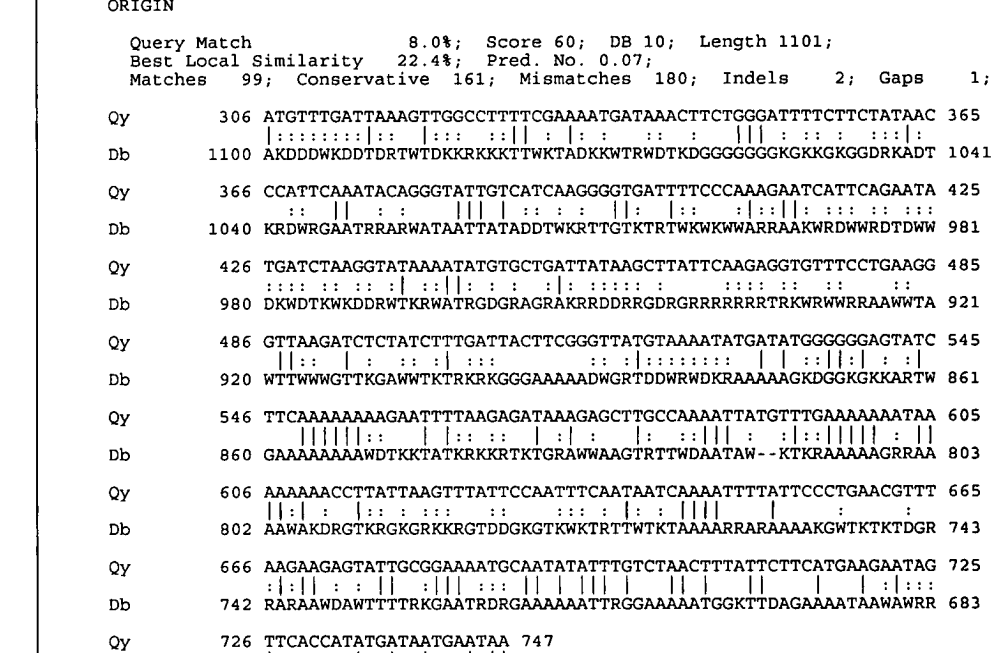
Query Match 8.0%; Score 60; DB 10; Length 1092;
Best Local Similarity 39.2%; Pred. No. 0.07;
Matches 123; Conservative 42; Mismatches 149; Indels 0; Gaps 0;



RESULT 7
CNS017KX/c
LOCUS CNS017KX 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37F10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL108171
VERSION AL108171.1 GI:5628475
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37F10"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : SP6"

ORIGIN

Query Match 8.0%; Score 60; DB 10; Length 1101;
Best Local Similarity 22.4%; Pred. No. 0.07;
Matches 99; Conservative 161; Mismatches 180; Indels 2; Gaps 1;



RESULT 8
CNS017ZQ
LOCUS CNS017ZQ 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37P08 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL108704
VERSION AL108704.1 GI:5629008
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"

```
                     /db_xref="taxon:7227"
                     /clone="BACN37P08"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             7.8%;  Score 58;  DB 10;  Length 1101;
  Best Local Similarity   34.3%;  Pred. No. 0.18;
  Matches  110;  Conservative   72;  Mismatches  138;  Indels    1;  Gaps    1;

Qy        399 GATTTTCCCAAAGAATCATTCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAA 458
              | |   : : ||: |  | :  | ||:|: :::| :    |  :|||: :   :  || :|
Db        779 GKTCMKASWAAMCAYACKYRAAAAAAWAWAHWMTKHTYCTTTTMATAYMYMTGRTMTAARA 838

Qy        459 GCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTA 518
               :: ::  |::| :| | :  :      ||   :: | ||| : |: |: :|:  : |||
Db        839 AMKRMDNKAMKATVTATAWWAWATTATATTTTAWWATWTATWWMTK-TYTYTYTTBTTTA 897

Qy        519 TGTAAAATATGATATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCT 578
              | : | |||| ||:          ||| |: :: :|||: | |:::: | :|| | :  |
Db        898 TWWTATATATTATWATAWTWATATATWTWWWWTWAAAWAATTWWWWWTATWTATATWTAT 957

Qy        579 TGCCAAAATTATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAAT 638
                |:|  | : |   | | |:||||| |:||   |||::  | |||    |:|| |::
Db        958 ATATAWATATYWATAAWATATAWATAAATAWAAATATATWWTATATATATATAWTTWAWW 1017

Qy        639 AATCAAAATTTTATTCCCTGAACGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTG 698
              ::: | :||| |:||   | |    |  | |: | :  :    | | :   ||||||  |
Db       1018 WWWTATWATTATWTTATATAATATATATATAWTATWTAWTTWTATATWTATATATATATA 1077

Qy        699 TCTAACTTTATTCTTCATGAA 719
              | |:  :|:::|  | :: |:
Db       1078 TATWTWWTWWWTWATWWWTAW 1098


RESULT 9
DU074282/c
LOCUS       DU074282                851 bp    DNA     linear   GSS 12-AUG-2005
DEFINITION  69279 Tomato HindIII BAC Library Lycopersicon esculentum genomic
            clone LE_HBa0128L08 5, genomic survey sequence.
ACCESSION   DU074282
VERSION     DU074282.1  GI:72510423
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 851)
  AUTHORS   Mueller,L.A., Buels,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J.,
            Van Eck,J. and Stack,S.
  TITLE     BAC end sequencing from three Solanum lycopersicon libraries
  JOURNAL   Unpublished (2005)
COMMENT     Other_GSSs: 69278
            Contact: Lukas Mueller
            Tanksley Lab, Dept. of Plant Breeding
            Cornell University
            251 Emerson Hall, Ithaca, NY 14853, USA
            Tel: 607-255-6557
            Fax: 607-255-6683
            Email: sgn-feedback@sgn.cornell.edu
            Insert Length: 134653  Std Error: 0.00
            Plate: 128  row: L  column: 8
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 49
            High quality sequence stop: 362.
FEATURES             Location/Qualifiers
     source          1. .851
                     /organism="Lycopersicon esculentum"
                     /mol_type="genomic DNA"
                     /cultivar="Heinz 1706"
                     /db_xref="taxon:4081"
                     /clone="LE_HBa0128L08"
                     /lab_host="E. coli"
                     /clone_lib="Tomato HindIII BAC Library"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             7.7%;  Score 57.8;  DB 10;  Length 851;
  Best Local Similarity   48.7%;  Pred. No. 0.19;
  Matches  152;  Conservative    0;  Mismatches  160;  Indels    0;  Gaps    0;

Qy        436 TATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCT 495
              || |||| ||        || || |  | | | || |       |    ||   |||| |
Db        721 TAAAAAAAATTATAAAATAATTAATTAAATAAAAAATAAAATAAAAAAAAATTAATTTAA 662

Qy        496 CTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGAGTATCTTCAAAAAAA 555
               ||   || | ||          ||   |||| |  | |         |   |  | ||| ||
Db        661 TTAATATTAAATATAATATAATAATAAAAAAAAAAAAAAAAAATAAAAATTTATAAAATAAT 602

Qy        556 AGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAACCTT 615
              | ||  | | |  |||| |    |   ||||  || |   || ||||||| | |||
Db        601 ATAAAATATAAATTTAAAAAAAATTTAAAAAAAATATANTAATAAAATAATATAAAATAA 542

Qy        616 ATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAACGTTTAAGAAGAGTA 675
              || |  |  |    |||| |||| | || || | ||   | |      ||  | | ||
Db        541 ATAATATAAAAAATAATTTAAATATTTAATATATAATATNTTAATTAAAAAAATATAATA 482

Qy        676 TTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAGAATAGTTCACCATAT 735
              ||    | ||    |||||| | |  ||  | ||| |  || || |||| ||    | |
Db        481 TTTAAAATAAAATTATATATATTTAAAAAATAATTATAAATTAATAATAATTATAAAAAA 422

Qy        736 GATAATGAATAA 747
                 |||  ||||
Db        421 ATNAATTTATAA 410


RESULT 10
CNS0039Q/c
LOCUS       CNS0039Q                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08M12 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063931
VERSION     AL063931.1  GI:4941788
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
```

```
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08M12"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             7.7%;  Score 57.4;  DB 10;  Length 1101;
  Best Local Similarity   32.0%;  Pred. No. 0.23;
  Matches  164;  Conservative   88;  Mismatches  261;  Indels    0;  Gaps    0;

Qy        232 AAAGGCGACTTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAA 291
              |||:  :  ||  |  :|:|   |||     :  |  : : :|  :  | | | |:|:|:|
Db       1080 AAARAGRGDTTKTTTKWTWAAAAAAAAAAAAKRAAAGRARARAGGDAAAAAAAADAKARA 1021

Qy        292 AATATCAAAGAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGA 351
              | |  |||     |:     :||| ||:|:|  |: : : :||  | |||       |
Db       1020 KAAAAAAAAATTGAAWAAAAAWTTAKAGKTKGKTTKAKGRGRAAAAAAAAAAAAAAAAAA 961

Qy        352 TTTTCTTCTATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAG 411
                : :: :|:|:    ||     |  |  : : ::  |  || |   : ||||   ::|
Db        960 AAAWTWWAWAWAWTTTTTTTTKTTTTAAARRRGKKAAAAAAAWGTTGRKTTTTTTTWWAA 901

Qy        412 AATCATTCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAG 471
              :   |   | |:| :     || |:|     |: |     :   :|::|  : :   :|: |
Db        900 RTAAAAAAAWAWTTWTKAAAAAAGKAAGTTAWTTWAATWWTAWAWWAATKGDGGGRARGG 841

Qy        472 GTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGAT 531
              | |:||  |::|     || |: | | |       ||  ::       | | :|:| | |
Db        840 GGGKTTTTTKWAAAAAAAAAAWTTTTTTTAAAATTTTTWWWAAAAATTTTGWAWTTTTAA 781

Qy        532 ATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATG 591
              |         : |    |||||||| ||    || | | ||: |  | :  ||: || |
Db        780 AAAATTTTTTWTTWAAAAAAAAAAAAAAAAAAAAAAAAAAAAAWWAAWTGRWAAARTTTTTT 721

Qy        592 TTTGAAAAAAATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTA 651
              |||| :::::   :::|||| |  : : :  :|    ||    || ||  |||| |
Db        720 TTTGGRRRRRGRGRRRAAAAAWTAKKGRGRRRDAAAARAAAARAAAAAAAAAAAAAAAAA 661

Qy        652 TTCCCTGAACGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTC 711
                | |   | :|| || :: |  : ::|||| :   ||| :|:::|     || |
Db        660 AAAAATTATTTTRWAAAAATWRAAAADAKRAAAASKTGSTATTKTDDTDATGTTTAAAAA 601

Qy        712 TTCATGAAGAATAGTTCACCATATGATAATGAA 744
                  |   :|::| |  |  | | |: :  :::
Db        600 AAATTTYYKAWWAATGRAAAAAAAGRARGRRRR 568


RESULT 11
CNS00YZS
LOCUS       CNS00YZS               1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN01M04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL097042
VERSION     AL097042.1  GI:5608653
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN01M04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             7.6%;  Score 57;  DB 10;  Length 1101;
  Best Local Similarity   38.8%;  Pred. No. 0.28;
  Matches  128;  Conservative   52;  Mismatches  142;  Indels    8;  Gaps    1;

Qy        426 TGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGG 485
              :: | :  | |: || ||:::|: |: :|  : : |||:  ||::|: :      |  ||
Db        721 KKKTTKTTGATHAAATATWWDTKKTKKBTTGRGRTTTAWAGAADDGRRRRGGGAGGGGGG 780

Qy        486 GTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGAGTATC 545
              |   :| | | | | | |:  |  :: ::::: :::    :|  | ||       | |:|
Db        781 GGGGRGGTTTATKTDTGTKTATKTKKKKKKKKKKKKKGGGKAAAAAATTTTTAAATTMTA 840

Qy        546 TTCAAAAAAAAGAATTT--------TAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAA 597
                |  |||||| | |||         | | | | | :    :|   ||||||     :  :
Db        841 ATTTAAAAAAAAATTTTTTAWTTAATTAAAAAAATWTTATKTWTTAAAATTTAAAAWWTW 900

Qy        598 AAAAATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCT 657
              ||:|:|:||:|||  || ||   ||||||  |||||  ||  |  |||| ||| |:
Db        901 AAWAWTWAAWAAAATTTTTTTTTATTTATTTWAATTTTAAATAAAAAAAATTTTTWAAAA 960

Qy        658 GAACGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATG 717
              |   :| || || | :|||     |     ||| : ||| | |   |||: | |: ||
Db        961 NANWNWTAAAAAATAAWATTTTTWNATTAAAAATTWTTTTTTTTTTTTTTWNTTTWAATT 1020

Qy        718 AAGAATAGTTCACCATATGATAATGAATAA 747
                |: :   |  :| :  |  : :|||:
Db       1021 TTTAWAWAAAWAWAWTTWTTTTTWAWATAW 1050


RESULT 12
CNS016CO
LOCUS       CNS016CO               1200 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15E04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106578
VERSION     AL106578.1  GI:5622626
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
```

```
     source          1. .1200
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15E04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             7.6%;  Score 56.6;  DB 10;  Length 1200;
  Best Local Similarity   32.4%;  Pred. No. 0.34;
  Matches  119;  Conservative   73;  Mismatches  175;  Indels    0;  Gaps    0;

Qy        283 ATATATAAAAATATCAAAGAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAA 342
              | | | ::::: : :    |    :  :||: |:||||||    :  |  :::|| ::||:
Db        802 AAAAAAMMMMMMMMGGGGKTRRRGCYTTDCTWAAAGTTTTTTHGGTTWWWWATAWWAAW 861

Qy        343 CTTCTGGGATTTTCTTCTATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATT 402
                :  | |    ||   : |:  ||||| :|  || | : | | |    | |
Db        862 AAWAAAAAAAWATTAAAAATTTTYTAWAAAAATAADGAWTAATTWTAACGAAAAATTTTA 921

Qy        403 TTCCCAAAGAATCATTCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTT 462
                :    ||  :: | : | || |: :| |      | || : : :  : | | | |  ::
Db        922 AWAATTAATTWWWAAYAAWAAAAWTWTTTTTWAATTTAATWTWTYTTWTAATTTTAAAWW 981

Qy        463 ATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTA 522
              |   |:    | ||:    |:    :| || :  |:  || | ||          :|| ::
Db        982 AAAAAWTTTTTTTTWAWATAWAWTAAWAWATAWAAAWTATTWAATATAAAAATAWATTWW 1041

Qy        523 AAATATGATATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCC 582
              |:|:|: :|::      |  :   | |:|||||| ||  :::| | | ||| |  ||
Db       1042 AWAWAWAWTWWAAWAAAAAAWAWATWAWAAAAAAWAAAAWWWAAAAAAAAAWAWWTTAAT 1101

Qy        583 AAAATTATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATC 642
              |:|:: :  | : :| |  :||||||: |  | : | : |||::|  |: || |    |
Db       1102 AWAWWAWAWTAWAWATATTWTAAAAAWTAWATAWATTWTTTTWWTWWAWATTTATATTTA 1161

Qy        643 AAAATTT 649
                |::|::
Db       1162 TAWWTWW 1168


RESULT 13
DN683496
LOCUS       DN683496                1367 bp    mRNA    linear   EST 30-MAR-2005
DEFINITION  CGX30-C12.x1d-t SHGC-CGX Gasterosteus aculeatus cDNA clone
            CGX30-C12 3', mRNA sequence.
ACCESSION   DN683496
VERSION     DN683496.1  GI:62029916
KEYWORDS    EST.
SOURCE      Gasterosteus aculeatus (three spined stickleback)
  ORGANISM  Gasterosteus aculeatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
            Gasterosteidae; Gasterosteus.
REFERENCE   1  (bases 1 to 1367)
  AUTHORS   Kingsley,D.M., Peichel,C., Balabahdra,S., Grimwood,J., Dickson,M.,
            Schmutz,J. and Myers,R.M.
  TITLE     Expressed sequence tags from Gasterosteus aculeatus
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Grimwood, Jane
            Stanford Human Genome Center
            Stanford University School of Medicine
            975 S California Ave, Palo Alto, CA 94304, USA
            Tel: 650 320 5917
            Fax: 650 320 5801
            Email: jane@shgc.stanford.edu
            Plate: 30
            High quality sequence start: 33
            High quality sequence stop: 989.
FEATURES             Location/Qualifiers
```

```
     source          1. .1367
                     /organism="Gasterosteus aculeatus"
                     /mol_type="mRNA"
                     /strain="Conner Creek sticklebacks, WA USA"
                     /db_xref="taxon:69293"
                     /clone="CGX30-C12"
                     /sex="mixed male and female"
                     /tissue_type="eyes"
                     /dev_stage="adult"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="SHGC-CGX"
                     /note="Vector: Express 1; Total and poly A+ RNA was
                     isolated from the indicated stickleback tissue, and a cDNA
                     library was constructed in the Express 1 plasmid vector by
                     Open Biosystems. First strand cDNA synthesis was primed
                     with an 54 bp linker primer containing an oligodT sequence
                     preceeded by a synthetic NotI site (first strand primer:
                     5'-GACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3'). Following
                     second strand synthesis, cDNAs were made blunt at the end
                     corresponding to the original 5 prime end of mRNA, and
                     cloned directionally into the NotI and EcoRV sites of
                     Express 1. Note that the EcoRV site is typically destroyed
                     in the blunt end cloning, leaving a junction of the form
                     'xxxATC'(where is ATC is the second half of the EcoRV
                     site, and xxx is derived from the cDNA sequence). A map of
                     the Express 1 vector is available at:
                     http://www.openbiosystems.com/cdna_library_construction_fa
                     q.php#8 The primary library was transformed and amplified
                     in DH10B (T1 phage resistant) bacteria. Clones available
                     from Open Biosystems:
                     http://www.openbiosystems.com/stickleback"
ORIGIN

  Query Match             7.6%;  Score 56.6;  DB 8;  Length 1367;
  Best Local Similarity   38.5%;  Pred. No. 0.34;
  Matches  182;  Conservative    0;  Mismatches  291;  Indels    0;  Gaps    0;

Qy        275 TTGGAGATATATATAAAAATATCAAAGAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAA 334
              || | | | | | ||||| | | |       | |||   |    ||     |||  |||
Db        282 TTTNAAAAAAAAAAAAAAAAAAAANANANNNANNNTNTTTTNNTTTTTTTNTTNNTTTTNAAA 341

Qy        335 ATGATAAACTTCTGGGATTTTCTTCTATAACCCATTCAAATACAGGGTATTGTCATCAAG 394
              ||  |||       ||  |     |  ||  |   |  ||  | |   | || |  | ||
Db        342 ATTTTAANTNNNGGGAAAAAGATAATAAGANAAAAAAAAGAAAATTTTTTTTTTTTAAAA 401

Qy        395 GGGTGATTTTCCCAAAGAATCATTCAGAATATGATCTAAGGTATAAAATATGTGCTGATT 454
              | | ||||         ||  |   | | |  |  |  |       |||  | | |
Db        402 AGTTTTTTTTNTTNTTTAAAAAAAAAGGAGGTTGTGGAANNNNNNAAATTGGGGGGGGGG 461

Qy        455 ATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACTTCGG 514
                  |      ||                 | | | | |||   |
Db        462 GGGGGGGGGAAAAAAAAAAAANANNNNNNNNNNNNNNNNTTTNTTTTTTTTTAAAAAAAAA 521

Qy        515 GTTATGTAAAATATGATATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAG 574
                |   ||    |  | |           | |  |||||||| | ||||       |
Db        522 AAAAAAAAATTNTTNNTTTNNNNNNNNNTTTTAAAAAAAAAAAATTTTTTTTTNNTTTTT 581

Qy        575 AGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTATTCCAATTT 634
                ||  ||||  ||     |||||||| ||||       | |  || |||  |
Db        582 TTTTTAAAAAAAAAATAANATAAAAAAAAAAAATTTTTTTAAAAAATTTTTTNNNNNNNA 641

Qy        635 CAATAATCAAAATTTTATTCCCTGAACGTTTAAGAAGAGTATTGCGGAAAATGCAATATA 694
              || ||  ||  ||  |     |  |     || || |         | ||     |  |
Db        642 AAAAAAAAAANTTTAAAAAAAATTTAAAAAAAAAAAAAAAANANNNNNANAAAAATTTTTT 701

Qy        695 TTTGTCTAACTTTATTCTTCATGAAGAATAGTTCACCATATGATAATGAATAA 747
              ||| | | |  || ||     |     || |   |  | || | ||  || ||
Db        702 TTTTTTTTAAATTTTTAAAAAANNNNAAAANNAAAAAAAATAAAAAAAAAAAA 754


RESULT 14
CW953272/c
LOCUS       CW953272                1047 bp    DNA     linear   GSS 20-DEC-2004
```

```
DEFINITION  TcB39.2_C07_SP6 Tribolium BAC library Tribolium castaneum genomic,
            genomic survey sequence.
ACCESSION   CW953272
VERSION     CW953272.1  GI:56736309
KEYWORDS    GSS.
SOURCE      Tribolium castaneum (red flour beetle)
  ORGANISM  Tribolium castaneum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
            Tenebrionidae; Tribolium.
REFERENCE   1  (bases 1 to 1047)
  AUTHORS   Savard,J. and Tautz,D.
  TITLE     Tribolium castaneum BAC-ends sequencing project
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Savard, J.
            Abteilung fur Evolutionsgenetik, AG Tautz
            Institut fur Genetik, Universitat zu Koln
            Weyertal 121, 50931 Koln, Germany
            Tel: 49 221 470 6911
            Fax: 49 221 470 5975
            Email: savard@uni-koeln.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1047
                     /organism="Tribolium castaneum"
                     /mol_type="genomic DNA"
                     /strain="GA-2"
                     /db_xref="taxon:7070"
                     /clone_lib="Tribolium BAC library"
                     /note="Vector: pBACe3.6; Site_1: EcoR1; Site_2: EcoR1;
                     Library constructed by Exelixis Inc."
ORIGIN

  Query Match             7.5%;  Score 55.8;  DB 10;  Length 1047;
  Best Local Similarity   43.6%;  Pred. No. 0.49;
  Matches  298;  Conservative    0;  Mismatches  382;  Indels    3;  Gaps    1;

Qy         27 TTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGTAGTTTATCAATTTTAAAAATAAA 86
              |||| ||| |  | |    |   | |    ||    ||  |  ||| ||||||  |   |
Db       1022 TTATTATATTATTTACTAATATTATATTTATTACTATATATATTCATTTTTAATATATAT 963

Qy         87 ACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACAGATGGAACGAATCGTGTCATTAG 146
              |    || | || |||||   |          | || | ||  |    ||    | ||
Db        962 AATAATTAATATCATTATCTATCTAAATTTATTTATAAATATATTCTATTTATTAATATC 903

Qy        147 TAGATTTACTAGTATGAATATTACACATGTTTATGAAAAAGATGAAGGGATATATGATGC 206
              |    || || ||| ||||||    || | | | | ||| ||  |      |  |
Db        902 AATTAATATTATTATAAATATTCATTATATATTTTATAAAAATATATATTATTTATAATT 843

Qy        207 GATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTAATACATTATTTAAACGCCGGCGA 266
              || |||||       |       | |||       || || |  || ||          |
Db        842 AATAAATAATAATATTAAAAAATCAAAAAATTTTGTATTAAAAAATATATTATAATATTA 783

Qy        267 TAGCGTAATTGGAGATATATATAAAAATATCAAAGAGCCATGTTTGATTAAAGTTGGCCT 326
              |      |  | |||  || || ||  || | |        | |  |  |   |  |
Db        782 TTTATAATCTTAATATAAATTTACAATCATTATAATTATCAATATTTTATACTATTACTA 723

Qy        327 TTTCGAAAATGATAAACTTCTGGGATTTTCTTCTATAACCCATTCAAATACAGGGTATTG 386
              | |  |   |  | | |   || | || |||   |  |   |||       |||
Db        722 TATATTATTAAACTTATATATACATTTATTTTATATTTTCTTTATTAATTATCCCTATCA 663

Qy        387 TCATCAAGGGGTGATTTTCCCAAAGAATCATTCAGAATATGATCTAAGGTATAAAATATG 446
              | ||||       |||| |  ||  | |    |  ||| | |    | |||| ||||
Db        662 TTATCATTATAATATTTACTTAATTATTTTATATTAATTTTACTATATATATAATATATT 603

Qy        447 TGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGAT 506
              |  | || |||  |   ||  |    | | |  | ||   |||            || ||
Db        602 TTTTTATAATATACAACTTATATCCATATATAATAAAACTTTATATAAATATAAATTTAT 543

Qy        507 TACTTCGGGTTATGTAAAATATGATATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAAG 566
              || ||      |  |||| |||  ||         | |    |||||||   || | |
Db        542 TAATTTAATCAAATTAAATTATTTTAATAATATTTTTTTAATAAAAAAAAATCTTCTTAT 483

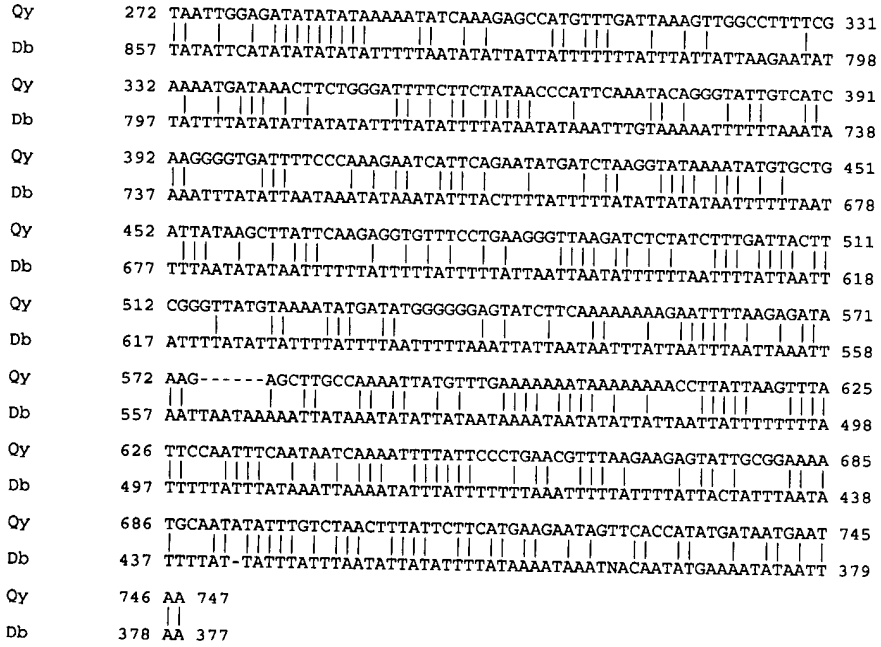
Qy        567 AGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTAT 626
              | | | | |   |  ||  |  ||   |    |  || | |  ||||| | |  |
Db        482 TATAATATATCATTATATTATACTTATTTTTATTTTTTTAATACATTTTATTTAATAAAA 423

Qy        627 TCCAATTTCAATAATCAAAATTTTATTCCCTGAACGTTTAAGAAGAGTATTGCGGAAAAT 686
              |   |||| | || |   ||| | |||    |      || |   | | | ||    |||||
Db        422 T---ATTTTATTATTATTAATATAATTATATATTAATTAATACATATTTTTAAATAAAAT 366

Qy        687 GCAATATATTTGTCTAACTTTAT 709
                ||   || | | |   |||||
Db        365 ATAAATCATATATATTTATTTAT 343


RESULT 15
CO389521/c
LOCUS       CO389521                878 bp    mRNA    linear   EST 30-JUN-2004
DEFINITION  AGENCOURT_26625198 NIH_MGC_253 Rattus norvegicus cDNA clone
            IMAGE:7304217 5', mRNA sequence.
ACCESSION   CO389521
VERSION     CO389521.1  GI:49497112
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 878)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical
            College of Wisconsin
             cDNA Library Preparation: Express Genomics
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM15340  row: h  column: 07
            High quality sequence stop: 120.
FEATURES             Location/Qualifiers
     source          1. .878
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="IMAGE:7304217"
                     /tissue_type="Ovary - Brown Norway Line 3 Age 8 weeks.
                     Tissues were snap-frozen and transfered in -70c. RNAse
                     free the entire procedure."
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_253"
                     /note="Organ: ovary; Vector: pExpress-1; Site_1: EcoRV;
                     Site_2: NotI; RNA obtained from female animals at 8 wk
                     old. Tissues were snap-frozen and kept at -80C for two
                     days before RNA extraction and purification (Tri-reagent
                     method). cDNA was primed using oligo-dT primer:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into
                     the EcoRV/NotI sites of pExpress-1. Size-selection >1.25kb
                     resulted in an average insert size of 1.5 kb. This primary
                     library is normalized (non-normalized primary library is
                     NIH_MGC_252) and was constructed by Express Genomics
                     (Frederick, MD). Note: this is a NIH_MGC library."
ORIGIN

  Query Match             7.4%;  Score 55;  DB 7;  Length 878;
  Best Local Similarity   47.5%;  Pred. No. 0.71;
  Matches  229;  Conservative    0;  Mismatches  246;  Indels    7;  Gaps    2;
```



Search completed: May 11, 2006, 05:00:30
Job time : 2395.49 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 20:34:42 ; Search time 104.196 Seconds
(without alignments)
12743.667 Million cell updates/sec

Title: US-09-423-093B-2_COPY_2011_2757
Perfect score: 747
Sequence: 1 atgaaaatatcagttataac.....caccatatgataatgaataa 747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	747	100.0	14187	3	US-09-453-702B-121	Sequence 121, App
c	2	747	100.0	14187	3	US-10-114-170-121	Sequence 121, App
c	3	51.8	6.9	50000	3	US-09-662-254B-23	Sequence 23, Appl
	4	49.8	6.7	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	5	46.8	6.3	640681	3	US-09-790-988-1	Sequence 1, Appli
c	6	46.4	6.2	605	2	US-08-358-918-36	Sequence 36, Appl
	7	46.4	6.2	25590	3	US-09-949-002-777	Sequence 777, App
c	8	46.2	6.2	314798	3	US-09-949-016-13539	Sequence 13539, A
	9	46.2	6.2	1664976	3	US-08-916-421B-1	Sequence 1, Appli
	10	46.2	6.2	1664976	3	US-09-692-570-1	Sequence 1, Appli
c	11	45.6	6.1	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	12	45.4	6.1	601	3	US-09-949-002-7550	Sequence 7550, Ap
	13	45.2	6.1	2082	3	US-09-440-325A-2	Sequence 2, Appli
	14	45.2	6.1	2082	3	US-09-846-996A-2	Sequence 2, Appli
	15	45.2	6.1	2082	3	US-10-246-085A-2	Sequence 2, Appli
	16	45	6.0	1472	3	US-08-956-171E-583	Sequence 583, App
	17	45	6.0	1472	3	US-08-781-986A-583	Sequence 583, App
c	18	45	6.0	32392	3	US-09-662-254B-27	Sequence 27, Appl
c	19	44.8	6.0	665	2	US-08-883-795A-36	Sequence 36, Appl
	20	44.6	6.0	50000	3	US-09-662-254B-23	Sequence 23, Appl
c	21	44.4	5.9	1230025	3	US-09-198-452A-1	Sequence 1, Appli
c	22	44.4	5.9	1230230	3	US-09-438-185A-1	Sequence 1, Appli
	23	44	5.9	187169	3	US-09-949-016-12776	Sequence 12776, A
	24	44	5.9	191569	3	US-09-949-016-15940	Sequence 15940, A
	25	44	5.9	640681	3	US-09-790-988-1	Sequence 1, Appli
c	26	43.6	5.8	50000	3	US-09-662-254B-25	Sequence 25, Appl
c	27	43.2	5.8	700	3	US-09-735-271-1038	Sequence 1038, Ap
	28	43.2	5.8	1885	3	US-09-662-254B-41	Sequence 41, Appl
	29	43.2	5.8	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	30	43.2	5.8	50000	3	US-09-662-254B-24	Sequence 24, Appl
	31	43.2	5.8	119153	3	US-09-949-016-12378	Sequence 12378, A
c	32	42.8	5.7	134987	3	US-09-949-016-15348	Sequence 15348, A
c	33	42.8	5.7	134987	3	US-09-949-016-15349	Sequence 15349, A
c	34	42.8	5.7	134987	3	US-09-949-016-15350	Sequence 15350, A
c	35	42.8	5.7	134987	3	US-09-949-016-15507	Sequence 15507, A
c	36	42.8	5.7	134987	3	US-09-949-016-15508	Sequence 15508, A
c	37	42.8	5.7	134987	3	US-09-949-016-15509	Sequence 15509, A
c	38	42.4	5.7	601	3	US-09-949-016-30531	Sequence 30531, A
c	39	42.4	5.7	601	3	US-09-949-016-37150	Sequence 37150, A
c	40	42.4	5.7	601	3	US-09-949-016-37164	Sequence 37164, A
c	41	42.4	5.7	601	3	US-09-949-016-145868	Sequence 145868,
c	42	42.4	5.7	601	3	US-09-949-016-146136	Sequence 146136,
c	43	42.4	5.7	601	3	US-09-949-016-146404	Sequence 146404,
c	44	42.4	5.7	601	3	US-09-949-002-7549	Sequence 7549, Ap
	45	42.4	5.7	205044	3	US-09-949-016-15851	Sequence 15851, A

ALIGNMENTS

```
RESULT 1
US-09-453-702B-121/c
; Sequence 121, Application US/09453702B
; Patent No. 6365723
;  GENERAL INFORMATION:
;       APPLICANT: Blattner, Frederick R.
;                  Burland,     Valerie
;                  Perna,  Nicole T.
;                  Plunkett, Guy
;                  Welch,  Rod
;       TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
;       NUMBER OF SEQUENCES: 265
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Quarles & Brady
;            STREET: 1 South Pinckney Street
;            CITY: Madison
;            STATE: WI
;            COUNTRY: US
;            ZIP: 53701-2113
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: Word Perfect 8.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/453,702B
;            FILING DATE: 03-Dec-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/110,955
;            FILING DATE: 04-DEC-1998
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Seay, Nicholas J.
;            REGISTRATION NUMBER: 27386
;            REFERENCE/DOCKET NUMBER: 960296.95017
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (608) 251-5000
;            TELEFAX: (608) 251-9166
;  INFORMATION FOR SEQ ID NO: 121:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 14187
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA  (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-09-453-702B-121
```

```
  Query Match             100.0%;  Score 747;  DB 3;  Length 14187;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-150;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11990 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 11931

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11930 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 11871

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11870 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 11811

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 11751

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 11691

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 11631

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 11571

Qy        421 GAATATGATCTAAGGTATAAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GAATATGATCTAAGGTATAAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 11511

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 11451

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 11391

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 11331

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 11271

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db      11270 AATAGTTCACCATATGATAATGAATAA 11244


RESULT 2
US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Patent No. 6855814
;   GENERAL INFORMATION:
;        APPLICANT: Blattner, Frederick R.
;                   Burland,     Valerie
;                   Perna,  Nicole T.
;                   Plunkett, Guy
;                   Welch,  Rod
;        TITLE OF INVENTION: No. 6855814el Sequences of E. coli O157
;        NUMBER OF SEQUENCES: 265
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Quarles & Brady
;             STREET: 1 South Pinckney Street
;             CITY: Madison
;             STATE: WI
;             COUNTRY: US
;             ZIP: 53701-2113
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Word Perfect 8.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/114,170
;             FILING DATE: 01-Apr-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/453,702
;             FILING DATE: 03-DEC-1999
;             APPLICATION NUMBER: 60/110,955
;             FILING DATE: 04-DEC-1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seay, Nicholas J.
;             REGISTRATION NUMBER: 27386
;             REFERENCE/DOCKET NUMBER: 960296.95017
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (608) 251-5000
;             TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 121:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14187
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA  (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121

  Query Match             100.0%;  Score 747;  DB 3;  Length 14187;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-150;
  Matches  747;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11990 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 11931

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11930 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 11871

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11870 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 11811

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 11751

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 11691

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 11631

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 11571

Qy        421 GAATATGATCTAAGGTATAAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GAATATGATCTAAGGTATAAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 11511

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      11510 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 11451

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 11391

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 11331

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 11271

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              ||||||||||||||||||||||||||
Db      11270 AATAGTTCACCATATGATAATGAATAA 11244
```

RESULT 3
US-09-662-254B-23/c
; Sequence 23, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-23

Query Match 6.9%; Score 51.8; DB 3; Length 50000;
Best Local Similarity 44.0%; Pred. No. 0.063;
Matches 267; Conservative 0; Mismatches 337; Indels 3; Gaps 1;

```
Qy         16 ATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGTAGTTTATCAATT 75
              |||| |   |  ||| || | | | |     | ||   | ||| |||| | | | |
Db      16934 ATAAATAAAATAAATATTATTAATAATGATCATAATAATTATATAAATAGTATTTTATTA 16875

Qy         76 TTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACAGATGGAACGAAT 135
              || ||  |||  |||||| |  |  |||| | ||   |  | ||    || || |||
Db      16874 AAAATAAACAAATATTTTGATAAAAAAATAGATAATAATGTATATAATTTTGACACAAAT 16815

Qy        136 CGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAAAAAGATGAAGGG 195
              | |   || || |  |   | |   |     | |     |||| ||  |||| |
Db      16814 ATTATATGTAATATAAATGAAAATGATATACAAATATTATATTATAAATTAGATTATTTT 16755

Qy        196 ATATATGATGCGATGAATAAGGG---CCGAATGTTGGCCAAAGGCGACTTAATACATTAT 252
              |||   ||    |  ||||         ||       ||| | || |  ||| | |||
Db      16754 ATACTCGAATATAAAAATATTTTATATAAAAATAATATAAAAAGTGATTATATAAAATAT 16695

Qy        253 TTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAAGAGCCATGTTTG 312
              | |       || | | |   | |||   | | |||||  | | |    | ||
Db      16694 ATTTATAATAATTGTATAGAATTATTAAATAACAAAATAAATAATATAAATAAGTATTAT 16635

Qy        313 ATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCTATAACCCATTCA 372
              ||||   |  | |  | |||||||  ||   ||    | | |    ||  |     |
Db      16634 ATTATTTTATGTATAATTAAAAATGAAAATAATCCTAAAATATTAAATAGTAATATATTT 16575

Qy        373 AATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCAGAATATGATCTA 432
```

```
              | ||      ||| | |   |      | ||    ||  | ||   |    | | ||
Db      16574 ATTAATATTGATTATGAATTATTATCTAATTATATAATTATTCCAACTATGTCTAATAGT 16515

Qy        433 AGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGA 492
              |  |||| |||| || | || || |      | |     |    | | ||    | |
Db      16514 AATAATAATATATATGATTATGATTATAAATGTTATATTGATAATGTTAATAACTACAAT 16455

Qy        493 TCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGAGTATCTTCAAAA 552
              |     | | |||  |      |  | ||||||        | | |||      | ||
Db      16454 CATAATAATATTATTGTTAACAAAAACTTTAAATATATAGATAAAGAATTATTCATTACA 16395

Qy        553 AAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAAC 612
              |||| |      |     ||||| ||      | |||| |   | ||  |||||||||
Db      16394 AAAAAATCGAATATTATAAAAGATCTATGTGGATTTATATACAATAATGATAAAAAAATT 16335

Qy        613 CTTATTA 619
              |  |||
Db      16334 TTATTTA 16328
```

RESULT 4
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 6.7%; Score 49.8; DB 3; Length 1141;
Best Local Similarity 10.7%; Pred. No. 0.11;
Matches 79; Conservative 264; Mismatches 386; Indels 6; Gaps 1;

```
Qy          8 TATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGTAGTT 67
              ::: ::::::|  :::::: ::::: ::   :::    :: |:::::   : : : :
Db        215 WWKNWSYDVTYYWWVWDDMCKRKVRRWVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMAYR 274

Qy         68 TATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACAGATG 127
              :|:       | |: ::|    :   ::   |  |:    |:::|:          |  :
Db        275 RWTMNNNNNNAKAMCKRAKYWGWNRABVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDH 334

Qy        128 GAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAAAAAG 187
              :::   | :|  :        :  :: :::|:::::::: |::  :    :: ::::    :
Db        335 KWWKWSAAMGVYWNNNNNNNWTYKKARHBARWDWVWHSAWKKWHANAAHYSRKKWTBYKR 394

Qy        188 ATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTAATAC 247
              | :     |  :::::   :: :::: |       | |       :::  ::::
Db        395 KTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYNNNNNGGRTYYGWTKNKKMWTYYKWKANN 454

Qy        248 ATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAAGAGCCAT 307
              ::: :::  | :     :::: :   : :     : :::|||: |:::::::  :|
Db        455 CKWRAWDHKTCTHNNTTWWKMKTYWNNCYWKSMTNGKSHRBAAAVYTWYMWWWRRYAHAN 514

Qy        308 GTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCTATAACCC 367
              :: :: |  ::: ::  ::    ||:  :  : :|:   :::|  ::    : :::
Db        515 NNNWDYWWKACTWYKYBVCSKWWNNYAAWYTKSSWNYTSRYYRWKTNNSWRWRSDTRSMG 574

Qy        368 ATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCAGAATATG 427
```

```
Db        575 RANNYARABHYGYKWNTRWWBWSHTWBHBRAGAAHYWMBMM------YBAKCHCMKAWYK 628

Qy        428 ATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGT 487

Db        629 AKKYAGAGGSNNNNNNNNNNNNNNNNATCARDDYYAASRWYAMANAKWYYYKBAANNAYY 688

Qy        488 TAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGAGTATCTT 547

Db        689 THANNWWGCWNNATDTRRTMWKNNNNNNAGTWKNNNNNNAKNASAAKNYAAAAVKAAKKH 748

Qy        548 CAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAA 607

Db        749 WRWANKWAMRGWHADAAABTTDKRNNGAYTKYTTTNNNNTYRGVVTNTAARDGWANNNNN 808

Qy        608 AAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAACGTTTAA 667

Db        809 NNNNNNNNNGWSDMWVTWWAYANYGTNNNNNNNNNNAYAWWTNKWYYTTDDRWRBAYTNN 868

Qy        668 GAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAGAATAGTT 727

Db        869 NNNNRMAYYGAYADDYAYYMSDTCDAWMKWDATKMNNATTYNRGTAWRTNNNNNNMTMKT 928

Qy        728 CACCATATGATAATG 742

Db        929 KYYBHAAWNNNNNNG 943
```

RESULT 5
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 6.3%; Score 46.8; DB 3; Length 640681;
Best Local Similarity 51.6%; Pred. No. 0.98;
Matches 131; Conservative 0; Mismatches 122; Indels 1; Gaps 1;

```
Qy         493 TCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGAGTATCTTCAAAA 552

Db      179592 TATTTTTTTGTATATACAATAGTCTATCATAAATATGATTAGATCGAAAAAAATACTTCA 179533

Qy         553 AAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAAC 612

Db      179532 AAAGTAATAGCATATAAAAAATCGGCCACTAAAAATACCTTTTTTATTAGTAAAAATTTT 179473

Qy         613 CTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAACGTTTAAGAAGA 672

Db      179472 TTTAGTATTATAAATATATTTTCAAAAAT-AATATTATATAGTACATTTAATTAAGCCAA 179414

Qy         673 GTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAGAATAGTTCACCA 732

Db      179413 TTAATGGATTAGGATATATACCTAAAGCTAACAGTATTATTCCTGAAATACATATTACTA 179354

Qy         733 TATGATAATGAATA 746

Db      179353 TACCAGAAGGAGTA 179340
```

RESULT 6
US-08-358-918-36/c
; Sequence 36, Application US/08358918
; Patent No. 5888774
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Erythropoietin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,918
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McDiarmid, Shona S.
; REGISTRATION NUMBER: P-38,798
; REFERENCE/DOCKET NUMBER: 7841-002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 605 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: Rh 32
US-08-358-918-36

Query Match 6.2%; Score 46.4; DB 2; Length 605;
Best Local Similarity 49.7%; Pred. No. 0.53;
Matches 147; Conservative 0; Mismatches 146; Indels 3; Gaps 1;

```
Qy         431 TAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAA 490

Db         434 TTAATTATAAAATATGTAATTATAAATACTTTATAAAATATGTAATTATAAAATATGTAA 375

Qy         491 GATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGAGTATCTTCAA 550

Db         374 TTATAAACATTTTAATTATAAAATATGTAATTATAAACATTTTAATTATAAAATATGTAA 315

Qy         551 AAAAAAGAATTTTA---AGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAA 607

Db         314 TTATAAACATTTTAATTATAAAATATGTAATTATAAACATTTTAATTATAAAATATTTAA 255

Qy         608 AAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAACGTTTAA 667

Db         254 TTATAAATATTTTAATTATAAAATATTTAATTATAAATATTTTAATTATAAAATATTTAA 195

Qy         668 GAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAGAAT 723

Db         194 TTATAAATATTTTAATTATAAAATATTTAATTATAAATATTTTAATTATAAAATAT 139
```

```
RESULT 7
US-09-949-002-777
; Sequence 777, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000790
;  CURRENT APPLICATION NUMBER: US/09/949,002
;  CURRENT FILING DATE:  2000-01-28
;  PRIOR APPLICATION NUMBER: 60/231,401
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 10823
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 777
;   LENGTH: 25590
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(25590)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-002-777

  Query Match             6.2%;  Score 46.4;  DB 3;  Length 25590;
  Best Local Similarity   46.6%;  Pred. No. 0.82;
  Matches  218;  Conservative    0;  Mismatches  246;  Indels    4;  Gaps    2;

Qy        272 TAATTGGAGATATATATAAAAATATCAAAGAGCCATGTTTGATTAAAGTTGGCCTTTTCG 331
              || || |  |||||||| | |   |   |      ||| | |  |       |   | |
Db       4153 TATTTTGTAATATATATTATATATTATTATTATATTGTATATTATTTTATATATTATATA 4212

Qy        332 AAAATGATAAACTTCTGGGATTTTCTTCTATAACCCATTCAAATACAGGGTATTGTCATC 391
              | |  || | |        |||  | ||| |   |||   |||     | |  |
Db       4213 TTATTATTATATTATATATTATTTAATATATTATATATTATTATATGTTATTTAATATAT 4272

Qy        392 AAGGGGTGATTTTCCCAAAGAATCATTCAGAATATGATCTAAGGTATAAAATATGTGCTG 451
              |    |||| |      |   || ||| |  |||| || || | |||| | ||| |  |
Db       4273 TATATATGATATATATTATATATTATTTAATATATTATATATGATATATATTATATATTA 4332

Qy        452 ATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGAT---TA 508
              |||  |  ||||  | ||  | | |  |  |   || |      ||| | ||||   ||
Db       4333 TTTAATATATTATATATGATATATATTATATATTATTTAATATATTATATATGATATATA 4392

Qy        509 CTTCGGGTTATGTAAAATATGATAT-GGGGGGAGTATCTTCAAAAAAAAGAATTTTAAGA 567
              |      |||| ||| |||| ||||        | ||| ||  | |  |    || |||
Db       4393 TTATATATTATTTAATATATTATATATTATTTAATATATTACATATGATATATATTATAT 4452

Qy        568 GATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTATT 627
              | || |  || |  |   ||| | | | | |  ||| | |   | ||| |  | ||||
Db       4453 ATTTAATATATTACATATGATATATATTATATATTTAATATATTACATATGATATATATT 4512

Qy        628 CCAATTTCAATAATCAAAATTTTATTCCCTGAACGTTTAAGAAGAGTATTGCGGAAAATG 687
              |  || ||||    | || | ||      |  | | |  |    |          | ||
Db       4513 ATATATTTAATATATTACATATGATATATATTATATATTTAATATATTACATATGATATA 4572

Qy        688 CAATATATTTGTCTAACTTTATTCTTCATGAAGAATAGTTCACCATAT 735
              | ||||| |   | | | |||| |  ||||   ||| |  |   |||
Db       4573 TATTATATATTATTTAATATATTATATATGAGATATATTATATATTAT 4620


RESULT 8
US-09-949-016-13539/c
; Sequence 13539, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13539
;   LENGTH: 314798
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(314798)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13539

  Query Match             6.2%;  Score 46.2;  DB 3;  Length 314798;
  Best Local Similarity   54.4%;  Pred. No. 1.2;
  Matches   93;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        327 TTTCGAAAATGATAAACTTCTGGGATTTTCTTCTATAACCCATTCAAATACAGGGTATTG 386
              ||   ||| | ||||| || |||  |||  | | || |    ||  | ||    ||| |
Db     137290 TTATAAAATTAATAAATTTGTGGTTTTTAATGCCATTAAAGTTTACAGTAATTTGTAATA 137231

Qy        387 TCATCAAGGGGTGATTTTCCCAAAGAATCATTCAGAATATGATCTAAGGTATAAAATATG 446
              || |||  | | | |    ||   |        |||||||| |||     ||| |    |
Db     137230 GCAGCAATAGATAACTAATACAGTCATCTTAGAAGAATATGGTCTCTTTAATAGATGTAG 137171

Qy        447 TGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCT 497
              | |  ||  ||  |||||||||  | ||  ||||| |  ||| || | |
Db     137170 AGATATTTTAAAATTTATTCAAGAAATATTCACTGAAAGTCTAAAATGTAT 137120


RESULT 9
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
```

; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g

US-08-916-421B-1

Query Match 6.2%; Score 46.2; DB 3; Length 1664976;
Best Local Similarity 52.3%; Pred. No. 1.5;
Matches 102; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```
Qy        415 CATTCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTG 474
              ||||    | | || ||     | | ||| || | |||    | |   ||    || ||
Db     229485 CATTATCCAAAGGAGCTTTTAAAGAGAATTTGAGGTGAAAGAATGAAAATCTCCGATGTT 229544

Qy        475 TTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATG 534
              ||  ||||   || |||     | || |   |    | |   |||||||| ||| | |
Db     229545 GTTGTTGAATTATTTAGAGAGGCAGCTATTTATCTACCAGAAGATGTAAAAAATGCTTTA 229604

Qy        535 GGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTT 594
              |  | || || |  |||| |'|| | |  || |   ||| |       |||   ||  |
Db     229605 GAAGAAGCATATAAAAAAGAAAGTAGTGAAATATCAAAAAACACATTAAAAGCAATCATA 229664

Qy        595 GAAAAAAATAAAAAA 609
              ||||| || |||| |
Db     229665 GAAAATAACAAAATA 229679
```

```
RESULT 10
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6797466
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275C1
;  CURRENT APPLICATION NUMBER: US/09/692,570
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  PRIOR APPLICATION NUMBER: US 08/916,421
;  PRIOR FILING DATE: 1997-08-22
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
```

; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g

Query Match 6.2%; Score 46.2; DB 3; Length 1664976;
Best Local Similarity 52.3%; Pred. No. 1.5;
Matches 102; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```
Qy      415 CATTCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTG 474
            ||||     | | || ||      | | ||| || | |||     | |   ||    || ||
Db   229485 CATTATCCAAAGGAGCTTTTAAAGAGAATTTGAGGTGAAAGAATGAAAATCTCCGATGTT 229544

Qy      475 TTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATG 534
            ||  ||||   || |||      | || |   |     | |   |||||||| ||| | |
Db   229545 GTTGTTGAATTATTTAGAGAGGCAGCTATTTATCTACCAGAAGATGTAAAAAATGCTTTA 229604

Qy      535 GGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTT 594
            |  | || || |  |||| |||| | |  || |   ||| |      |||   || |
Db   229605 GAAGAAGCATATAAAAAAGAAAGTAGTGAAATATCAAAAAACACATTAAAAGCAATCATA 229664

Qy      595 GAAAAAAATAAAAAA 609
            ||||| || |||| |
Db   229665 GAAAATAACAAAATA 229679
```

RESULT 11
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 6.1%; Score 45.6; DB 3; Length 1141;
Best Local Similarity 10.0%; Pred. No. 0.85;
Matches 69; Conservative 274; Mismatches 344; Indels 3; Gaps 1;

```
Qy       50 AAAAAACTTTAAGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTG 109
            ||::|:     :  |:|||:| : ::|::  : ::::  ::||| : :|| | :|
Db      782 AARMARTCNNYMHAAVTTTHTDWCYKTWMNTWYWDMMTTMBTTTTRNMTTSTNMTNNNNN 723

Qy      110 ATGGCGGCTCTACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTA 169
            |       :::|:  :|   : | :   :  :         : ::   : ::: ||:
Db      722 NMWACTNNNNNNMWKAYYAHATNNWGCWWNNTDARRTNNTTVMRRRWMTNTKTRWYSTTR 663

Qy      170 CACATGTTTATGAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGG 229
             :: || |                 :  | |:    :  ::|:   :::::  ::|  | :
Db      662 RHHYTGATNNNNNNNNNNNNNNNSCCTCTRMMTMRWTMKGDGMTVRKKVKWRDTTCTYV 603

Qy      230 CCAAAGGCGACTTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATA 289
            ::|:: : ::  ::   : | |      |:::  : ::::   :::   ::  |  :  :
Db      602 DVWADSWVWWYANWMRCRDVTYTRNNTYCKSYAHSYWYWSNNAMWY---RRYSARNWSSM 546

Qy      290 AAAATATCAAAGAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGG 349
            |:: |:  :::: :: :   | :: ::     :  :::: :::|::  |: ::    ::
Db      545 ARWTTRNNWWMSGBVRMRWAGTMWWRHWNNNNTDTRYYWWWKRWARBTTTVYDSMCNAKS 486

Qy      350 GATTTTCTTCTATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAA 409
             :      :::::|   :   :::: :: |     :   |: ::     :: :
Db      485 MWRGNNWRAMKMWWAANNDAGAMDHWTYWMGNNTMWMRRAWKMMNMAWCRRAYCCNNNNN 426
```

```
Qy        410 AGAATCATTCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAG 469
Db        425 RACVWHKHKMWRWTWKYMWKAACNNNNBKAMYMRVAWMMYSRDTTNTDWMMWTSDWBWHW 366

Qy        470 AGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATG 529
Db        365 YTVDYTMMRAWNNNNNNNWRBCKTTSWMWWMDHMNTHCTYGNNTWGSAYBMAAMSMWAAG 306

Qy        530 ATATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTA 589
Db        305 ASNBVTYNWCWRMTYMGKTMTNNNNNNKAWYYRTKTVAWCNNRYYYDTAVWTBKRNYKYC 246

Qy        590 TGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTT 649
Db        245 YAYBWYYBMYMGKHHWBWWRRABHRSWNMWWVKCRNKYMVSWHYHAMRYBKWABAVGCNN 186

Qy        650 TATTCCCTGAACGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTAT 709
Db        185 NWKDRMAHHHWCATNNNMMWWWYAYMHHMHKKGKAAWTNNKTABRDDHBAHVKTYWYWRY 126

Qy        710 TCTTCATGAAGAATAGTTCACCATATGATA 739
Db        125 DYWCAMCWMNAKAKVRTAMKHMWYYTDRYV 96


RESULT 12
US-09-949-002-7550/c
; Sequence 7550, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000790
;  CURRENT APPLICATION NUMBER: US/09/949,002
;  CURRENT FILING DATE:  2000-01-28
;  PRIOR APPLICATION NUMBER: 60/231,401
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 10823
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7550
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-002-7550

  Query Match             6.1%;  Score 45.4;  DB 3;  Length 601;
  Best Local Similarity   46.6%;  Pred. No. 0.87;
  Matches  214;  Conservative    0;  Mismatches  241;  Indels    4;  Gaps    2;

Qy        272 TAATTGGAGATATATATAAAAATATCAAAGAGCCATGTTTGATTAAAGTTGGCCTTTTCG 331
Db        464 TATTTTGTAATATATATTATATATTATTATTATATTGTATATTATTTTATATATTATATA 405

Qy        332 AAAATGATAAACTTCTGGGATTTTCTTCTATAACCCATTCAAATACAGGGTATTGTCATC 391
Db        404 TTATTATTATATTATATATTATTTAATATATTATATATTATTATATGTTATTTAATATAT 345

Qy        392 AAGGGGTGATTTTCCCAAAGAATCATTCAGAATATGATCTAAGGTATAAAATATGTGCTG 451
Db        344 TATATATGATATATATTATATATTATTTAATATATTATATATGMTATATATTATATATTA 285

Qy        452 ATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGAT---TA 508
Db        284 TTTAATATATTATATATGATATATATTATATATTATTTAATATATTATATATGATATATA 225

Qy        509 CTTCGGGTTATGTAAAATATGATAT-GGGGGGAGTATCTTCAAAAAAAAGAATTTTAAGA 567
Db        224 TTATATATTATTTAATATATTATATATTATTTAATATATTACATATGATATATATTATAT 165

Qy        568 GATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTATT 627
Db        164 ATTTAATATATTACATATGATATATATTATATATTTAATATATTACATATGATATATATT 105

Qy        628 CCAATTTCAATAATCAAAATTTTATTCCCTGAACGTTTAAGAAGAGTATTGCGGAAAAATG 687
Db        104 ATATATTTAATATATTACATATGATATATATTATATATTTAATATATTACATATGATATA 45

Qy        688 CAATATATTTGTCTAACTTTATTCTTCATGAAGAATAGT 726
Db         44 TATTATATATTATTTAATATATTATATATGAGATATATT 6


RESULT 13
US-09-440-325A-2
; Sequence 2, Application US/09440325A
; Patent No. 6280994
; GENERAL INFORMATION:
;  APPLICANT:  Sheppard, Paul O.
;  TITLE OF INVENTION:  Zac1: A Human Metalloenzyme
;  FILE REFERENCE:  98-79
;  CURRENT APPLICATION NUMBER: US/09/440,325A
;  CURRENT FILING DATE:  1999-11-15
;  NUMBER OF SEQ ID NOS:  3
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2082
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: This degenerate sequence encodes the amino acid
;   OTHER INFORMATION: sequence of SEQ ID NO:1.
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(2082)
;   OTHER INFORMATION: n = A,T,C or G
US-09-440-325A-2

  Query Match             6.1%;  Score 45.2;  DB 3;  Length 2082;
  Best Local Similarity   28.0%;  Pred. No. 1.1;
  Matches   97;  Conservative   57;  Mismatches  193;  Indels    0;  Gaps    0;

Qy        118 TCTACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTT 177
Db        220 TTYATGGARGCNACNTGGAAYTAYGTNACNAAYATHACNMGNAARAAYCARGARGARATG 279

Qy        178 TATGAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGC 237
Db        280 ATGAARGAYATGGARMGNWSNCARTTYATGATHTAYTTYGGNACNCARGCNCAYYTNTTY 339

Qy        238 GACTTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATC 297
Db        340 AARGTNACNCARTTYAARGAYCCNGAYGTNAAYGGNATGYTNWSNAARYTNCARAAYATH 399

Qy        298 AAAGAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCT 357
Db        400 GAYAARGCNGCNYTNWSNAARGAYGARYTNMGNGARTAYAAYGARYTNYTNGCNNNNYTN 459

Qy        358 TCTATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCAT 417
Db        460 GARATGACNTAYWSNATGGCNCARGTNTGYYTNAAYGARGGNCCNTGYYTNWSNYTNGAR 519

Qy        418 TCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTAT 464
Db        520 WSNGARYTNGARGTNATGGCNACNWSNMGNGAYAARGARGARYTNYT 566


RESULT 14
US-09-846-996A-2
; Sequence 2, Application US/09846996A
; Patent No. 6524836
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  TITLE OF INVENTION: Zace1: A Human Metalloenzyme
;  FILE REFERENCE: 98-79D1
```

```
;  CURRENT APPLICATION NUMBER: US/09/846,996A
;  CURRENT FILING DATE:  2001-05-01
;  PRIOR APPLICATION NUMBER: US 09/440,325
;  PRIOR FILING DATE: 1999-11-15
;  PRIOR APPLICATION NUMBER: US 60/109,783
;  PRIOR FILING DATE: 1998-11-25
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2082
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: This degenerate sequence encodes the amino acid
;   OTHER INFORMATION: sequence of SEQ ID NO:1.
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(2082)
;   OTHER INFORMATION: n = A,T,C or G
US-09-846-996A-2

  Query Match             6.1%;  Score 45.2;  DB 3;  Length 2082;
  Best Local Similarity   28.0%;  Pred. No. 1.1;
  Matches   97;  Conservative   57;  Mismatches  193;  Indels    0;  Gaps    0;

Qy        118 TCTACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTT 177
              | :|  || |  ||       : : | |  |   |:|| :| | :||:     : |  |
Db        220 TTYATGGARGCNACNTGGAAYTAYGTNACNAAYATHACNMGNAARAAYCARGARGARATG 279

Qy        178 TATGAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGC 237
                |: |    ||: | :   |   :  |||  |:   ||       :    :     :
Db        280 ATGAARGAYATGGARMGNWSNCARTTYATGATHTAYTTYGGNACNCARGCNCAYYTNTTY 339

Qy        238 GACTTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATC 297
              |  | |  || | :  : |: | | :|  | :|  ||         | ::   |:||:||:
Db        340 AARGTNACNCARTTYAARGAYCCNGAYGTNAAYGGNATGYTNWSNAARYTNCARAAYATH 399

Qy        298 AAAGAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCT 357
              |  |: |    :| :  ||:| :|  :|    ||: |: |: |::| :| |     :
Db        400 GAYAARGCNGCNYTNWSNAARGAYGARYTNMGNGARTAYAAYGARYTNYTNGCNNNNYTN 459

Qy        358 TCTATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCAT 417
                || ||  |::: |   |   :   ||::  | | :|    | ::  :      |
Db        460 GARATGACNTAYWSNATGGCNCARGTNTGYYTNAAYGARGGNCCNTGYYTNWSNYTNGAR 519

Qy        418 TCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTAT 464
              :: ||::  ||  | | |   |  :   | | : |  |  |::|  |
Db        520 WSNGARYTNGARGTNATGGCNACNWSNMGNGAYAARGARGARYTNYT 566


RESULT 15
US-10-246-085A-2
; Sequence 2, Application US/10246085A
; Patent No. 6878810
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  TITLE OF INVENTION: Zace1: A Human Metalloenzyme
;  FILE REFERENCE: 98-79D2
;  CURRENT APPLICATION NUMBER: US/10/246,085A
;  CURRENT FILING DATE:  2002-12-10
;  PRIOR APPLICATION NUMBER: US 09/846,996
;  PRIOR FILING DATE: 2001-05-01
;  PRIOR APPLICATION NUMBER: US 09/440,325
;  PRIOR FILING DATE: 1999-11-15
;  PRIOR APPLICATION NUMBER: US 60/109,783
;  PRIOR FILING DATE: 1998-11-25
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 2082
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: This degenerate sequence encodes the amino acid
;   OTHER INFORMATION: sequence of SEQ ID NO:1.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(2082)
;   OTHER INFORMATION: n = A,T,C or G
US-10-246-085A-2

  Query Match             6.1%;  Score 45.2;  DB 3;  Length 2082;
  Best Local Similarity   28.0%;  Pred. No. 1.1;
  Matches   97;  Conservative   57;  Mismatches  193;  Indels    0;  Gaps    0;

Qy        118 TCTACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTT 177
              | :|  || |  ||       : : | |  |   |:|| :| | :||:     : |  |
Db        220 TTYATGGARGCNACNTGGAAYTAYGTNACNAAYATHACNMGNAARAAYCARGARGARATG 279

Qy        178 TATGAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGC 237
                |: |    ||: | :   |   :  |||  |:   ||       :    :     :
Db        280 ATGAARGAYATGGARMGNWSNCARTTYATGATHTAYTTYGGNACNCARGCNCAYYTNTTY 339

Qy        238 GACTTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATC 297
              |  | |  || | :  : |: | | :|  | :|  ||         | ::   |:||:||:
Db        340 AARGTNACNCARTTYAARGAYCCNGAYGTNAAYGGNATGYTNWSNAARYTNCARAAYATH 399

Qy        298 AAAGAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCT 357
              |  |: |    :| :  ||:| :|  :|    ||: |: |: |::| :| |     :
Db        400 GAYAARGCNGCNYTNWSNAARGAYGARYTNMGNGARTAYAAYGARYTNYTNGCNNNNYTN 459

Qy        358 TCTATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCAT 417
                || ||  |::: |   |   :   ||::  | | :|    | ::  :      |
Db        460 GARATGACNTAYWSNATGGCNCARGTNTGYYTNAAYGARGGNCCNTGYYTNWSNYTNGAR 519

Qy        418 TCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTAT 464
              :: ||::  ||  | | |   |  :   | | : |  |  |::|  |
Db        520 WSNGARYTNGARGTNATGGCNACNWSNMGNGAYAARGARGARYTNYT 566


Search completed: May 11, 2006, 05:22:54
Job time : 113.196 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:01:11 ; Search time 493.042 Seconds
(without alignments)
12528.801 Million cell updates/sec

Title: US-09-423-093B-2_COPY_2011_2757
Perfect score: 747
Sequence: 1 atgaaaatatcagttataac.....caccatatgataatgaataa 747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	747	100.0	14024	3	US-09-423-093A-2	Sequence 2, Appli
c	2	747	100.0	14187	5	US-10-114-170-121	Sequence 121, App
	3	73.4	9.8	744	7	US-10-282-122A-41909	Sequence 41909, A
	4	55.2	7.4	14516	3	US-09-423-093A-1	Sequence 1, Appli
	5	52.2	7.0	3673778	6	US-10-312-841-1	Sequence 1, Appli
c	6	51.8	6.9	50000	8	US-10-706-635-23	Sequence 23, Appl
	7	51.6	6.9	1011	7	US-10-282-122A-15651	Sequence 15651, A
c	8	51.4	6.9	3089	9	US-10-954-778-160	Sequence 160, App
c	9	51.4	6.9	4985	5	US-10-094-240-10	Sequence 10, Appl
c	10	51.4	6.9	4985	6	US-10-056-405-10	Sequence 10, Appl
c	11	51	6.8	29993	8	US-10-482-823-3	Sequence 3, Appli
c	12	51	6.8	29993	8	US-10-482-823-5	Sequence 5, Appli
c	13	51	6.8	38342	7	US-10-221-714A-472	Sequence 472, App
c	14	50.8	6.8	5647	6	US-10-311-455-1539	Sequence 1539, Ap
c	15	50.8	6.8	5647	7	US-10-221-613-281	Sequence 281, App
	16	50.8	6.8	8056	8	US-10-473-126-386	Sequence 386, App
c	17	50.6	6.8	10189	6	US-10-311-455-2134	Sequence 2134, Ap
	18	50.6	6.8	158001	7	US-10-211-179-11	GENERAL INFORMATI
c	19	50.4	6.7	8056	8	US-10-473-126-386	Sequence 386, App
	20	49.6	6.6	6992	3	US-09-767-041-9	Sequence 9, Appli
	21	49.6	6.6	6992	3	US-09-767-041-29	Sequence 29, Appl
c	22	49	6.6	647	8	US-10-767-795-263	Sequence 263, App
	23	49	6.6	8056	8	US-10-473-126-240	Sequence 240, App
c	24	48.8	6.5	34722	7	US-10-322-281-700	Sequence 700, App
	25	48.6	6.5	333	8	US-10-357-930-12856	Sequence 12856, A
c	26	48.2	6.5	8170	6	US-10-240-453-132	Sequence 132, App
c	27	48	6.4	5987	6	US-10-311-455-1536	Sequence 1536, Ap
c	28	48	6.4	5987	7	US-10-433-793-132	Sequence 132, App
	29	48	6.4	12669	7	US-10-221-613-412	Sequence 412, App
	30	47.8	6.4	1014	7	US-10-398-221-855	Sequence 855, App
	31	47.8	6.4	1014	7	US-10-398-221-2769	Sequence 2769, Ap
c	32	47.8	6.4	1501	8	US-10-473-126-328	Sequence 328, App
	33	47.8	6.4	15479	7	US-10-257-166-46	Sequence 46, Appl
	34	47.6	6.4	2703	7	US-10-282-122A-40657	Sequence 40657, A
	35	47.4	6.3	930	7	US-10-282-122A-18332	Sequence 18332, A
c	36	47.2	6.3	3222	4	US-09-925-065A-674172	Sequence 674172,
	37	47.2	6.3	16217	6	US-10-311-455-598	Sequence 598, App
	38	47.2	6.3	18154	6	US-10-311-455-228	Sequence 228, App
	39	47	6.3	5999	6	US-10-311-455-531	Sequence 531, App
	40	47	6.3	5999	7	US-10-257-166-41	Sequence 41, Appl
	41	47	6.3	12507	6	US-10-311-455-272	Sequence 272, App
c	42	47	6.3	14950	6	US-10-311-455-1230	Sequence 1230, Ap
c	43	47	6.3	3673778	6	US-10-312-841-2	Sequence 2, Appli
	44	46.8	6.3	7810	5	US-10-239-676-141	Sequence 141, App
	45	46.8	6.3	7810	6	US-10-240-453-155	Sequence 155, App

ALIGNMENTS

RESULT 1
US-09-423-093A-2
; Sequence 2, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 14024
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-423-093A-2

Query Match 100.0%; Score 747; DB 3; Length 14024;
Best Local Similarity 100.0%; Pred. No. 2.4e-139;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2011 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 2070

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2071 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 2130

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2131 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 2190

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2191 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 2250

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2251 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 2310
```

```
Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2311 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 2370

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2371 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 2430

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2431 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 2490

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2491 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 2550

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2551 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 2610

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2611 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 2670

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2671 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 2730

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db       2731 AATAGTTCACCATATGATAATGAATAA 2757
```

RESULT 2
US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14187
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121

Query Match 100.0%; Score 747; DB 5; Length 14187;
Best Local Similarity 100.0%; Pred. No. 2.4e-139;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11990 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 11931

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11930 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 11871

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11870 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 11811

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 11751

Qy        241 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATATAAAAATATCAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 TTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAA 11691

Qy        301 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 GAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCT 11631

Qy        361 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 ATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCA 11571

Qy        421 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCT 11511

Qy        481 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 GAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGA 11451

Qy        541 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 GTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAA 11391

Qy        601 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 AATAAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAA 11331

Qy        661 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 CGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAG 11271

Qy        721 AATAGTTCACCATATGATAATGAATAA 747
              |||||||||||||||||||||||||||
Db      11270 AATAGTTCACCATATGATAATGAATAA 11244
```

RESULT 3
US-10-282-122A-41909
; Sequence 41909, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:

```
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41909
;   LENGTH: 744
;   TYPE: DNA
;   ORGANISM: Yersinia pestis
US-10-282-122A-41909

  Query Match             9.8%;  Score 73.4;  DB 7;  Length 744;
  Best Local Similarity   54.7%;  Pred. No. 7.6e-05;
  Matches  146;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||| |||    |||| |||| || ||||||| ||    |    |      | ||  ||
Db          1 ATGAAGATAAGTATTATTACTGCTACATATAATAGTGAGAGAACAATTTCGGATACCCTA 60

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
                |  |||    |    ||          | |||||  |||||| |||||||| |  ||
Db         61 TCTTCTTTGGAGAAACAAACCTATCTGGATGTTGAGTATATTATTGTTGATGGTGCATCA 120

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              |  |||   ||      ||| ||||||     | |||       | ||    | ||  |
Db        121 AAGGATAATACATTAACTGTTATTAGTCAGAATTCTACGCGAGTCACTAAGATTATTAGT 180

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGAC 240
              ||    ||| |||| || ||||||||  |||||||  ||    |   |||||  | ||| |||
Db        181 GAGTCTGATCAAGGTATTTATGATGCTTTGAATAAAGGGATAGATTTGGCAACAGGTGAC 240

Qy        241 TTAATACATTATTTAAACGCCGGCGAT 267
                | ||   || |||| |  | |  |||
Db        241 ATTATTGGTTTTTTACATTCTGATGAT 267


RESULT 4
US-09-423-093A-1
; Sequence 1, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
;  APPLICANT:  Reeves, Peter Richard et al
;  TITLE OF INVENTION:  Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
;  TITLE OF INVENTION:  Thereof
;  FILE REFERENCE:  749670/PAS
;  CURRENT APPLICATION NUMBER: US/09/423,093A
;  CURRENT FILING DATE:  1999-11-01
;  PRIOR APPLICATION NUMBER:  PO6545
;  PRIOR FILING DATE:  1997-05-01
;  PRIOR APPLICATION NUMBER:  PO8162
;  PRIOR FILING DATE:  1997-07-22
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 14516
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-09-423-093A-1

  Query Match             7.4%;  Score 55.2;  DB 3;  Length 14516;
  Best Local Similarity   50.8%;  Pred. No. 0.98;
  Matches  132;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy          8 TATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGTAGTT 67
              | ||| | || | ||  |||||||||  ||      |   |||  ||   |||
Db      10989 TTTCAATAATCATTGCAACTTATAATTCTGAACTTGATATAGCTAAGTGTTTGCAATCGG 11048

Qy         68 TATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACAGATG 127
              || | | |  |       ||   | |||| || || ||| | ||||| || ||| | |||
Db      11049 TAACTAATCAATCTTATAAGAATATTGAAATCATAATAATGGATGGAGGATCTTCTGATA 11108

Qy        128 GAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAAAAAG 187
              ||||  |  | |     |     ||||            || | |   ||||  || ||||
Db      11109 AAACGCTTGATATTGCAAAATCGTTTAAAGACGACCGAATAAAAATAGTTTCAGAGAAAG 11168

Qy        188 ATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTAATAC 247
              ||   || || ||||||||   |||||| |         ||  |||  || || |   ||
Db      11169 ATCGTGGAATTTATGATGCCTGGAATAAAGCAGTTGATTTATCCATTGGTGATTGGGTAG 11228

Qy        248 ATTATTTAAACGCCGGCGAT 267
                | | |     | |  |||
Db      11229 CATTTATTGGTTCAGATGAT 11248


RESULT 5
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3294164)
US-10-312-841-1

  Query Match             7.0%;  Score 52.2;  DB 6;  Length 3673778;
  Best Local Similarity   46.5%;  Pred. No. 28;
  Matches  168;  Conservative    0;  Mismatches  193;  Indels    0;  Gaps    0;

Qy        272 TAATTGGAGATATATATAAAAATATCAAAGAGCCATGTTTGATTAAAGTTGGCCTTTTCG 331
              |||||   |  || | |  | ||||   |     |      |  ||| |||      |
```

```
Db     340888 TAATTTTTGTAATTTTTGTAGATATTTTATTAGGAATAAATAGAAAATTTGAATAGATTA 340947

Qy        332 AAAATGATAAACTTCTGGGATTTTCTTCTATAACCCATTCAAATACAGGGTATTGTCATC 391

Db     340948 ATAATGAGTAATAAGATTGAATTAGTAATTAAAAGTGTTTTAATAAAGAAAAGTTTAGGA 341007

Qy        392 AAGGGGTGATTTTCCCAAAGAATCATTCAGAATATGATCTAAGGTATAAAATATGTGCTG 451

Db     341008 TTAGATGGTTTTTTTGTTAAATTTTATTATATATATAAAGAAAAATTAATATTAATATTT 341067

Qy        452 ATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACTT 511

Db     341068 TTTAAAATATTTAAAAAAAATTGAAGAGGAAGGAATTATTTTTAATTTATTTTATGAAGT 341127

Qy        512 CGGGTTATGTAAAATATGATATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATA 571

Db     341128 TTGTATTGTTTTGATATTAAAAGTAGAGAAGAATATAAAAAAAAAGAAAATTATAGGTTA 341187

Qy        572 AAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTATTCCAA 631

Db     341188 ATTTTTTTAGTGAAAATATATATAAAAATTTTGAATAAAATTTTAGTAAATTGAATTTAA 341247

Qy        632 T 632

Db     341248 T 341248
```

RESULT 6
US-10-706-635-23/c
; Sequence 23, Application US/10706635
; Publication No. US20050014263A1
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XCZ1
; CURRENT APPLICATION NUMBER: US/10/706,635
; CURRENT FILING DATE: 2003-11-12
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 09/662,254
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-10-706-635-23

Query Match 6.9%; Score 51.8; DB 8; Length 50000;
Best Local Similarity 44.0%; Pred. No. 7.3;
Matches 267; Conservative 0; Mismatches 337; Indels 3; Gaps 1;

```
Qy         16 ATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGTAGTTTATCAATT 75

Db      16934 ATAAATAAAATAAATATTATTAATAATGATCATAATAATTATATAAATAGTATTTTATTA 16875

Qy         76 TTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACAGATGGAACGAAT 135

Db      16874 AAAATAAACAAATATTTTGATAAAAAAATAGATAATAATGTATATAATTTTGACACAAAT 16815

Qy        136 CGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAAAAAGATGAAGGG 195

Db      16814 ATTATATGTAATATAAATGAAAATGATATACAAATATTATATTATAAATTAGATTATTTT 16755

Qy        196 ATATATGATGCGATGAATAAGGG---CCGAATGTTGGCCAAAGGCGACTTAATACATTAT 252

Db      16754 ATACTCGAATATAAAAATATTTTATATAAAAATAATATAAAAAGTGATTATATAAAATAT 16695

Qy        253 TTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAAAGAGCCATGTTTG 312

Db      16694 ATTTATAATAATTGTATAGAATTATTAAATAACAAAATAAATAATATAAATAAGTATTAT 16635

Qy        313 ATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTCTATAACCCATTCA 372

Db      16634 ATTATTTTATGTATAATTAAAAATGAAAATAATCCTAAAATATTAAATAGTAATATATTT 16575

Qy        373 AATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTCAGAATATGATCTA 432

Db      16574 ATTAATATTGATTATGAATTATTATCTAATTATATAATTATTCCAACTATGTCTAATAGT 16515

Qy        433 AGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGA 492

Db      16514 AATAATAATATATATGATTATGATTATAAATGTTATATTGATAATGTTAATAACTACAAT 16455

Qy        493 TCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGAGTATCTTCAAAA 552

Db      16454 CATAATAATATTATTGTTAACAAAAACTTTAAATATATAGATAAAGAATTATTCATTACA 16395

Qy        553 AAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAAC 612

Db      16394 AAAAAATCGAATATTATAAAAGATCTATGTGGATTTATATACAATAATGATAAAAAAATT 16335

Qy        613 CTTATTA 619

Db      16334 TTATTTA 16328
```

RESULT 7
US-10-282-122A-15651
; Sequence 15651, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1

```
; SEQ ID NO 15651
;   LENGTH: 1011
;   TYPE: DNA
;   ORGANISM: Clostridium acetobutylicum
US-10-282-122A-15651

  Query Match             6.9%;  Score 51.6;  DB 7;  Length 1011;
  Best Local Similarity   51.0%;  Pred. No. 2;
  Matches  151;  Conservative    0;  Mismatches  139;  Indels    6;  Gaps    1;

Qy          4 AAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGT 63
              ||| ||||| |  |||        ||||||| ||        |||| | || || ||
Db         16 AAAGTATCAATAGTAATGCCAGTATATAATAGTGAGAGGTATTTAGCAGAAGCTATAGAA 75

Qy         64 AGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACA 123
              ||| | | |  |    |      ||   |||||||| |||| |||||||||| || |||||
Db         76 AGTATTTTAGATCAGACCTATAATGATTTTGAGTTTATAATAGTTGATGATGGTTCTACT 135

Qy        124 GATGGAACGAATCGTGTCATTAGTAGATTTACTAGT------ATGAATATTACACATGTT 177
              |||| | |  ||    | ||   ||| |    ||| |      |  | ||||       ||
Db        136 GATGAATCTTATAACATAATATCTAGCTACGCTAATAAAGACAACAGAATTATTGTGATT 195

Qy        178 TATGAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGC 237
              |    | || ||   ||  |    ||| |  | ||| | ||   ||    | || | |||
Db        196 TCAAGAGAACATAGGGGACTTGTCGATTCATTAAATGAAGGTATAAACATAGCTAGAGGA 255

Qy        238 GACTTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAA 293
               | |  |||       |  | || |  ||||     |||    |||  |    ||||
Db        256 AAATACATAGCAAGAATGGATGCGGATGATATATCCATTAATAATAGAATCGAAAA 311


RESULT 8
US-10-954-778-160/c
; Sequence 160, Application US/10954778
; Publication No. US20050153368A1
; GENERAL INFORMATION:
;  APPLICANT: ZWIEBEL, LAURENCE J.
;  TITLE OF INVENTION: METHOD OF IDENTIFYING CHEMICAL AGENTS WHICH STIMULATE
;  TITLE OF INVENTION:  ODORANT RECEPTORS OF SENSORY NEURONS
;  FILE REFERENCE: N1125
;  CURRENT APPLICATION NUMBER: US/10/954,778
;  CURRENT FILING DATE:  2004-09-30
;  PRIOR APPLICATION NUMBER: 10/056,405
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: 60/264,649
;  PRIOR FILING DATE: 2001-01-26
;  NUMBER OF SEQ ID NOS: 248
;  SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 160
;   LENGTH: 3089
;   TYPE: DNA
;   ORGANISM: Anopheles gambiae
US-10-954-778-160

  Query Match             6.9%;  Score 51.4;  DB 9;  Length 3089;
  Best Local Similarity   43.8%;  Pred. No. 3.2;
  Matches  314;  Conservative    0;  Mismatches  401;  Indels    2;  Gaps    2;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              || | |||| | | |||||  |  |    |  |       |    |  |  ||   ||
Db       2508 ATAAGAATAATAATAATAACAATAGCAATAATAAGAATAATAATAACAATAGCAATAATA 2449

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              |  |   ||  |||  ||| ||||| |     |     || || | |     ||        |
Db       2448 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAAAAAAAAATAATTATAAT 2389

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              |   || | |  |||  |   | || ||      | || ||| |||| ||   ||  | ||
Db       2388 AAGAATAGTAATAATAATAATAATAATATTAAAAATAATATTAATAATAATAATAATAAT 2329

Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAA-GGGCCGAATGTTGGCCAAAGGCGA 239
               | ||  || |    |  || ||   | |||||       |||  |    | |
Db       2328 AATAATAATAAGAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2269

Qy        240 CTTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATCAA 299
              | ||||| | |  |||        |  |  ||||     |||   |||| ||||    |
Db       2268 CATAATAAAATTATTATTATGATAATAATGTATAATAATGAATAACAATAATAATAATTA 2209

Qy        300 AGAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAAATGATAAACTTCTGGGATTTTCTTC 359
              | |   ||  | | | || |     |     | ||| ||||   |    | | |  |
Db       2208 ATAATAATAAATTAATTAATTAATAATAAATAATAATAATAATTATTATTGTTATAATAA 2149

Qy        360 TATAACCCATTCAAATACAGGGTATTGTCATCAAGGGGTGATTTTCCCAAAGAATCATTC 419
              ||  |   ||   ||||       ||  | ||||       | |       || ||| ||
Db       2148 TAATAATAATAATAATAATAATAATAATAATCATAATTATAATAATAATAATAATAATAA 2089

Qy        420 AGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCC 479
              |  || ||   |  |||| |     |  | || ||||    | | |  |       |
Db       2088 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAAAGTAATAATTATAATAA 2029

Qy        480 TGA-AGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGG 538
              | | ||  ||| |    ||    | || | |         ||| |      ||
Db       2028 TAATAGTAATAATAATAATAATAATAATAATATTAATAATAATAATAATAATAATAATAA 1969

Qy        539 GAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAA 598
              | ||     || ||  | |||  ||| |   | |     |    | ||| ||  |   ||
Db       1968 TAATAATAATAATAATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 1909

Qy        599 AAAATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTG 658
              ||  | || ||   | || |   | ||   |||   |||||| ||   | ||||
Db       1908 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAATTCTGTATTTTGGT 1849

Qy        659 AACGTTTAAGAAGAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCA 715
               |  | |  ||     || |  ||||||   ||| ||| ||| |   ||    |||
Db       1848 TCCTGTAACTAAATTGGTTACTAAAAATGGTTTATCTTTCTCTTAAAATAAAAATCA 1792


RESULT 9
US-10-094-240-10/c
; Sequence 10, Application US/10094240
; Publication No. US20030082637A1
; GENERAL INFORMATION:
;  APPLICANT: ZWIEBEL, LAURENCE J.
;  TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
;  FILE REFERENCE: N8289
;  CURRENT APPLICATION NUMBER: US/10/094,240
;  CURRENT FILING DATE:  2001-03-08
;  PRIOR APPLICATION NUMBER: 10/056,405
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: 60/264,649
;  PRIOR FILING DATE: 2001-01-26
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 4985
;   TYPE: DNA
;   ORGANISM: Anopheles gambiae
US-10-094-240-10

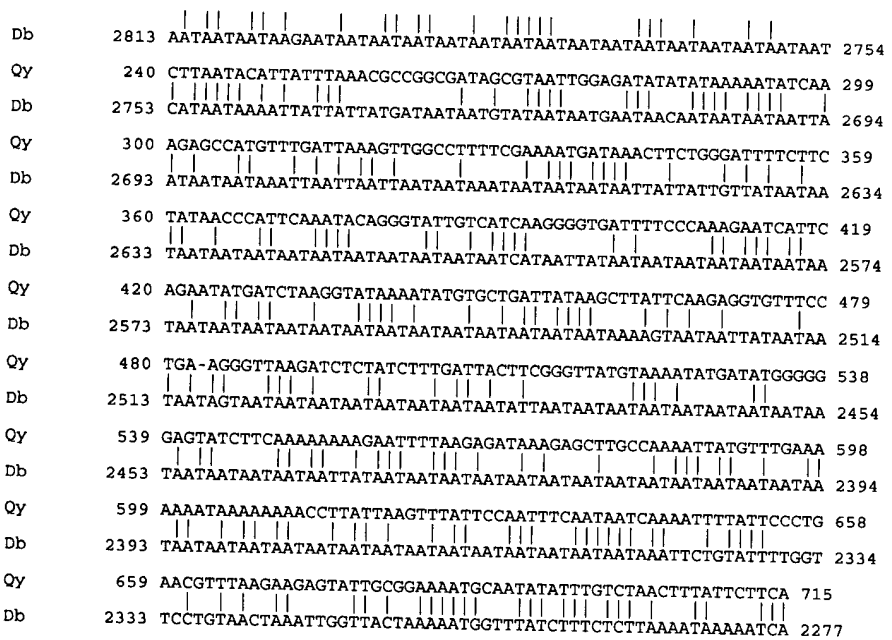
  Query Match             6.9%;  Score 51.4;  DB 5;  Length 4985;
  Best Local Similarity   43.8%;  Pred. No. 3.9;
  Matches  314;  Conservative    0;  Mismatches  401;  Indels    2;  Gaps    2;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              || | |||| | | |||||  |  |    |  |       |    |  |  ||   ||
Db       2993 ATAAGAATAATAATAATAACAATAGCAATAATAAGAATAATAATAACAATAGCAATAATA 2934

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              |  |   ||  |||  ||| ||||| |     |     || || | |     ||        |
Db       2933 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAAAAAAAAATAATTATAAT 2874

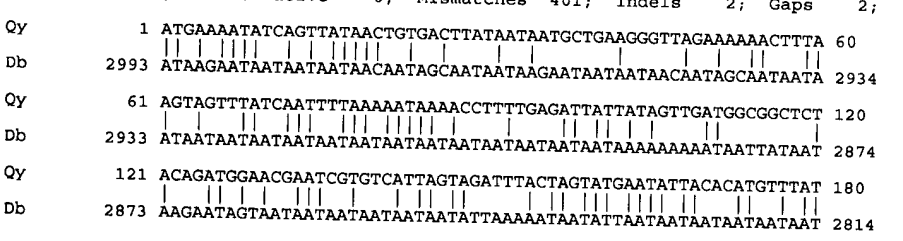
Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTAT 180
              |   || | |  |||  |   | || ||      | || ||| |||| ||   ||  | ||
Db       2873 AAGAATAGTAATAATAATAATAATAATATTAAAAATAATATTAATAATAATAATAATAAT 2814

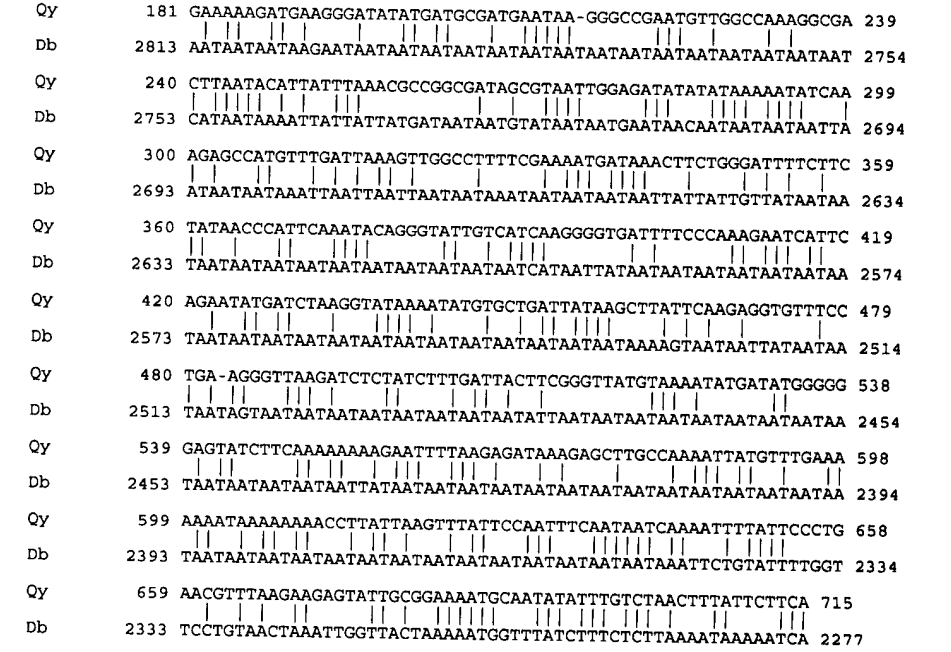
Qy        181 GAAAAAGATGAAGGGATATATGATGCGATGAATAA-GGGCCGAATGTTGGCCAAAGGCGA 239
```



RESULT 10
US-10-056-405-10/c
; Sequence 10, Application US/10056405
; Publication No. US20030166013A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: MOSQUITO OLFACTORY GENES, POLYPEPTIDES, AND METHODS OF
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: N7841
; CURRENT APPLICATION NUMBER: US/10/056,405
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-056-405-10

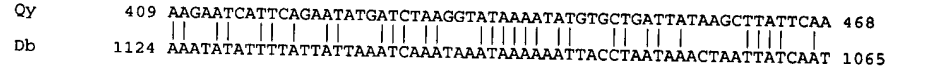
Query Match 6.9%; Score 51.4; DB 6; Length 4985;
Best Local Similarity 43.8%; Pred. No. 3.9;
Matches 314; Conservative 0; Mismatches 401; Indels 2; Gaps 2;





RESULT 11
US-10-482-823-3/c
; Sequence 3, Application US/10482823
; Publication No. US20050003463A1
; GENERAL INFORMATION:
; APPLICANT: Adorjan, Peter
; APPLICANT: Burger, Matthias
; APPLICANT: Maier, Sabine
; APPLICANT: Lesche, Ralph
; APPLICANT: Cottrell, Susan
; APPLICANT: De Vos, Theo
; TITLE OF INVENTION: Method and nucleic acids for the analysis of colorectal cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE: 47675-55/P119US
; CURRENT APPLICATION NUMBER: US/10/482,823
; CURRENT FILING DATE: 2003-12-30
; PRIOR APPLICATION NUMBER: PCT/EP03/01457
; PRIOR FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: EP 02004551.4
; PRIOR FILING DATE: 2002-02-27
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 3
; LENGTH: 29993
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-482-823-3

Query Match 6.8%; Score 51; DB 8; Length 29993;
Best Local Similarity 50.6%; Pred. No. 8.8;
Matches 123; Conservative 0; Mismatches 120; Indels 0; Gaps 0;



```
Qy        469 GAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATAT 528
              | ||| || ||   |||| || |     ||  || | |     |||| |||||  |
Db       1064 ATTATATTTTCTAAAAAATTAAAATATTCCCATTCAATAAATAATACTTATCTAAAAATT 1005

Qy        529 GATATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATT 588
               |||         |       | |||| | || || ||   | |   |   |       |
Db       1004 TCTATCTATATTATCCTAAAACAAAACAAAAATTAAAATAAAATCAATTATATACCTAAA 945

Qy        589 ATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATT 648
              |  | | ||||||| |||||||| ||| | ||    |||     |  |||||  | |  |
Db        944 AAATATAAAAAAAAAAAAAAAAAATCTTTTAAAAAAAATTTTTCCTAAAATAACTATACAT 885

Qy        649 TTA 651
              |||
Db        884 TTA 882


RESULT 12
US-10-482-823-5/c
; Sequence 5, Application US/10482823
; Publication No. US20050003463A1
; GENERAL INFORMATION:
;  APPLICANT: Adorjan, Peter
;  APPLICANT:  Burger, Matthias
;  APPLICANT:  Maier, Sabine
;  APPLICANT:  Lesche, Ralph
;  APPLICANT:  Cottrell, Susan
;  APPLICANT:  De Vos, Theo
;  TITLE OF INVENTION: Method and nucleic acids for the analysis of colorectal cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE: 47675-55/P119US
;  CURRENT APPLICATION NUMBER: US/10/482,823
;  CURRENT FILING DATE:  2003-12-30
;  PRIOR APPLICATION NUMBER: PCT/EP03/01457
;  PRIOR FILING DATE: 2003-02-13
;  PRIOR APPLICATION NUMBER: EP 02004551.4
;  PRIOR FILING DATE: 2002-02-27
;  NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 5
;   LENGTH: 29993
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-482-823-5

  Query Match             6.8%;  Score 51;  DB 8;  Length 29993;
  Best Local Similarity   50.6%;  Pred. No. 8.8;
  Matches  123;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy        409 AAGAATCATTCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAA 468
              ||  ||  || |  ||   ||| ||   |||||| ||   || ||| |     ||||  |
Db       1124 AAATATATTTTATTATTAAATCAAATAAATAAAAAATTACCTAATAAACTAATTATCAAT 1065

Qy        469 GAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATAT 528
              | ||| || ||   |||| || |     ||  || | |     |||| |||||  |
Db       1064 ATTATATTTTCTAAAAAATTAAAATATTCCCATTCAATAAATAATACTTATCTAAAAATT 1005

Qy        529 GATATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATT 588
               |||         |       | |||| | || || ||   | |   |   |       |
Db       1004 TCTATCTATATTATCCTAAAACAAAACAAAAATTAAAATAAAATCAATTATATACCTAAA 945

Qy        589 ATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATT 648
              |  | | ||||||| |||||||| ||| | ||    |||     |  |||||  | |  |
Db        944 AAATATAAAAAAAAAAAAAAAAAATCTTTTAAAAAAAATTTTTCCTAAAATAACTATACAT 885

Qy        649 TTA 651
              |||
Db        884 TTA 882


RESULT 13
US-10-221-714A-472/c
; Sequence 472, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 472
;   LENGTH: 38342
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-472

  Query Match             6.8%;  Score 51;  DB 7;  Length 38342;
  Best Local Similarity   50.6%;  Pred. No. 9.6;
  Matches  123;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy        409 AAGAATCATTCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAA 468
              ||  ||  || |  ||   ||| ||   |||||| ||   || ||| |     ||||  |
Db       8139 AAATATATTTTATTATTAAATCAAATAAATAAAAAATTACCTAATAAACTAATTATCAAT 8080

Qy        469 GAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATAT 528
              | ||| || ||   |||| || |     ||  || | |     |||| |||||  |
Db       8079 ATTATATTTTCTAAAAAATTAAAATATTCCCATTCAATAAATAATACTTATCTAAAAATT 8020

Qy        529 GATATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATT 588
               |||         |       | |||| | || || ||   | |   |   |       |
Db       8019 TCTATCTATATTATCCTAAAACAAAACAAAAATTAAAATAAAATCAATTATATACCTAAA 7960

Qy        589 ATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATT 648
              |  | | ||||||| |||||||| ||| | ||    |||     |  |||||  | |  |
Db       7959 AAATATAAAAAAAAAAAAAAAAAATCTTTTAAAAAAAATTTTTCCTAAAATAACTATACAT 7900

Qy        649 TTA 651
              |||
Db       7899 TTA 7897


RESULT 14
US-10-311-455-1539/c
; Sequence 1539, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
```

```
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1539
;   LENGTH: 5647
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1539

  Query Match             6.8%;  Score 50.8;  DB 6;  Length 5647;
  Best Local Similarity   54.2%;  Pred. No. 5.3;
  Matches  103;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        551 AAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAA 610
              ||  || |  || || | | ||| |   | ||||  | || ||| ||||| || || |||
Db        961 AATCAATATATTCAATATAAAAAAAAAATCCCAATCTAATTTTTCAAAAATATTAACAAA 902

Qy        611 ACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAACGTTTAAGAA 670
              || |||  ||  |  ||  | |  ||  |||| || |     | ||  | || ||
Db        901 CTAATACTAAAATTTATATAAAATAACCAAAAAAAACTTAAAAAAATAAAAAATCAAAAA 842

Qy        671 GAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAGAATAGTTCAC 730
                ||     |||||  ||||  ||  | |||  ||||| |   | || ||   |  ||
Db        841 ATAAATCCTATAAAATATAATAATTTATTATAAAATTATTATAATTAAAAAAATATAAAC 782

Qy        731 CATATGATAA 740
                || | ||
Db        781 TTAATAAAAA 772


RESULT 15
US-10-221-613-281/c
; Sequence 281, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, ALexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
;  FILE REFERENCE: 5013.1004
;  CURRENT APPLICATION NUMBER: US/10/221,613
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: PCT/EP01/02945
;    DE 10013847.00
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-03-15
;    2000-03-15
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 281
;   LENGTH: 5647
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-281

  Query Match             6.8%;  Score 50.8;  DB 7;  Length 5647;
  Best Local Similarity   54.2%;  Pred. No. 5.3;
  Matches  103;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        551 AAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAA 610
              ||  || |  || || | | ||| |   | ||||  | || ||| ||||| || || |||
Db        961 AATCAATATATTCAATATAAAAAAAAAATCCCAATCTAATTTTTCAAAAATATTAACAAA 902

Qy        611 ACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAACGTTTAAGAA 670
              || |||  ||  |  ||  | |  ||  |||| || |     | ||  | || ||
Db        901 CTAATACTAAAATTTATATAAAATAACCAAAAAAAACTTAAAAAAATAAAAAATCAAAAA 842

Qy        671 GAGTATTGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAGAATAGTTCAC 730
                ||     |||||  ||||  ||  | |||  ||||| |   | || ||   |  ||
Db        841 ATAAATCCTATAAAATATAATAATTTATTATAAAATTATTATAATTAAAAAAATATAAAC 782

Qy        731 CATATGATAA 740
                || | ||
Db        781 TTAATAAAAA 772


Search completed: May 11, 2006, 07:08:24
Job time : 500.042 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:23:58 ; Search time 386.013 Seconds
(without alignments)
7892.621 Million cell updates/sec

Title: US-09-423-093B-2_COPY_2011_2757
Perfect score: 747
Sequence: 1 atgaaaatatcagttataac.....caccatatgataatgaataa 747

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	71.2	9.5	762	18	US-11-079-463-2002	Sequence 2002, Ap
	2	69.6	9.3	903	18	US-11-079-463-566	Sequence 566, App
	3	59.2	7.9	786	18	US-11-079-463-1845	Sequence 1845, Ap
	4	54.4	7.3	966	17	US-11-098-686-8881	Sequence 8881, Ap
c	5	54.4	7.3	194553	17	US-11-098-686-8738	Sequence 8738, Ap
	6	52.4	7.0	783	18	US-11-079-463-2777	Sequence 2777, Ap
	7	49.4	6.6	891	18	US-11-079-463-838	Sequence 838, App
	8	47.8	6.4	2944528	18	US-11-045-004-1	Sequence 1, Appli
c	9	47.2	6.3	3222	7	US-09-925-065A-674172	Sequence 674172,
	10	46.4	6.2	990	12	US-10-301-480-568794	Sequence 568794,
	11	46.4	6.2	990	12	US-10-301-480-1182203	Sequence 1182203,
	12	46	6.2	579	12	US-10-301-480-293547	Sequence 293547,
	13	46	6.2	579	12	US-10-301-480-906956	Sequence 906956,
	14	46	6.2	587	7	US-09-925-065A-205659	Sequence 205659,
	15	46	6.2	1082144	17	US-11-117-187-211	Sequence 211, App
c	16	45	6.0	528	12	US-10-301-480-255605	Sequence 255605,
c	17	45	6.0	528	12	US-10-301-480-255606	Sequence 255606,
c	18	45	6.0	528	12	US-10-301-480-869014	Sequence 869014,
c	19	45	6.0	528	12	US-10-301-480-869015	Sequence 869015,
c	20	45	6.0	562	7	US-09-925-065A-163097	Sequence 163097,
c	21	45	6.0	562	7	US-09-925-065A-163098	Sequence 163098,
	22	44.8	6.0	579	12	US-10-301-480-293548	Sequence 293548,
	23	44.8	6.0	579	12	US-10-301-480-293549	Sequence 293549,
	24	44.8	6.0	579	12	US-10-301-480-906957	Sequence 906957,
	25	44.8	6.0	579	12	US-10-301-480-906958	Sequence 906958,
	26	44.8	6.0	587	7	US-09-925-065A-205660	Sequence 205660,
	27	44.8	6.0	587	7	US-09-925-065A-205661	Sequence 205661,
	28	44.8	6.0	1146	17	US-11-098-686-9829	Sequence 9829, Ap
c	29	44.8	6.0	1457619	17	US-11-098-686-8739	Sequence 8739, Ap
	30	44.4	5.9	747	12	US-10-301-480-585203	Sequence 585203,
	31	44.4	5.9	747	12	US-10-301-480-1198612	Sequence 1198612,
	32	43.8	5.9	6317	10	US-10-240-708-12	Sequence 12, Appl
c	33	43.6	5.8	641	7	US-09-925-065A-800820	Sequence 800820,
	34	43.6	5.8	753	18	US-11-079-463-543	Sequence 543, App
c	35	43.6	5.8	173602	17	US-11-121-086-25	Sequence 25, Appl
c	36	43.4	5.8	1174	10	US-10-750-185-32772	Sequence 32772, A
c	37	43.4	5.8	1174	10	US-10-750-623-32772	Sequence 32772, A
	38	43.4	5.8	1993	10	US-10-750-185-43425	Sequence 43425, A
	39	43.4	5.8	1993	10	US-10-750-623-43425	Sequence 43425, A
	40	43.4	5.8	158692	17	US-11-121-086-30	Sequence 30, Appl
	41	43.4	5.8	1457619	17	US-11-098-686-8739	Sequence 8739, Ap
	42	43	5.8	568	7	US-09-925-065A-951167	Sequence 951167,
	43	42.8	5.7	651	7	US-09-925-065A-728316	Sequence 728316,
	44	42.6	5.7	465	7	US-09-925-065A-364802	Sequence 364802,
	45	42.6	5.7	1731	10	US-10-750-185-31100	Sequence 31100, A

ALIGNMENTS

RESULT 1
US-11-079-463-2002
; Sequence 2002, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAC
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 2002
; LENGTH: 762
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-2002

Query Match 9.5%; Score 71.2; DB 18; Length 762;
Best Local Similarity 54.6%; Pred. No. 0.0036;
Matches 142; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
Qy          8 TATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGTAGTT 67
              | ||  |||| || ||  | ||||||  |   ||| |  | ||  | ||  | |  |||
Db         47 TTTCGATTATTACAGTATCATATAATGCTAAAGAAAGTATTGAGGATACGATTAACAGTA 106

Qy         68 TATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACAGATG 127
              |    | |  ||        |  ||||||   |||||| |||||||| ||   ||||||||
Db        107 TTGTTAATCAAACGTATTCAAATTTTGAATATATTATTGTTGATGGAGGTAGTACAGATG 166

Qy        128 GAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAAAAAG 187
              ||||     | | ||  | | || |    ||  | |          |   |||   ||
Db        167 GAACATTAGATATAATACGAAAATATGAGGATAAAATTTCATATTGGATAAGTGAGCCAG 226
```

```
Qy        188 ATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTAATAC 247
              || | || || |||||||| |||||||| ||      |     ||||| || |  ||
Db        227 ATAATGGTATTTATGATGCAATGAATAAAGGATTGGCTATTTGTAAAGGTGAATGGATTA 286

Qy        248 ATTATTTAAACGCCGGCGAT 267
              ||| | | ||   ||| |||
Db        287 ATTTTATGAATAGCGGGGAT 306
```

RESULT 2
US-11-079-463-566
; Sequence 566, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 566
; LENGTH: 903
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-566

Query Match 9.3%; Score 69.6; DB 18; Length 903;
Best Local Similarity 54.2%; Pred. No. 0.0068;
Matches 141; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

```
Qy          8 TATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGTAGTT 67
              | || ||||| || |||  ||||||| | ||   | | | ||    ||| |   ||||
Db         89 TGTCTGTTATTACGGTGTGTTATAATGCTACTACTGTGATTGAGGCTACTATTCTTAGTA 148

Qy         68 TATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACAGATG 127
              |          ||| |      | | ||   ||| ||  | ||||| ||   ||||||||
Db        149 TCATTGGGCAAACATATTCCAATATCGAATATATCATTATAGATGGTGGTAGTACAGATG 208

Qy        128 GAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAAAAAG 187
              |||| |     ||||| |  | || |   |  |  |          |||   |||   |
Db        209 GAACAATAGAAGTCATCAAGAAATATGAAAAGAAAATATCCTATTGGGTTAGCGAACCGG 268

Qy        188 ATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTAATAC 247
              || ||||| || |||||| || |||||||| ||      |      | |  || ||  | || |
Db        269 ATAAAGGTATTTATGACGCTATGAATAAAGGTATTGTAAAATCGACTGGAGAATGGATCC 328

Qy        248 ATTATTTAAACGCCGGCGAT 267
              ||| | |||| || || |||
Db        329 ATTTTCTAAATGCAGGAGAT 348

RESULT 3
US-11-079-463-1845
; Sequence 1845, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 1845
; LENGTH: 786
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-1845

Query Match 7.9%; Score 59.2; DB 18; Length 786;
Best Local Similarity 54.0%; Pred. No. 0.4;
Matches 149; Conservative 0; Mismatches 118; Indels 9; Gaps 1;

```
Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||| || ||  |||||||||| |||| ||||| ||   ||     |  | || |
Db         22 ATGAAGATTTCGATTATAACTGTCACTTTTAATAGTGAAAAAACTATTAATGATACCATT 81

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
                | | |  ||   |          | ||||   ||| ||| |||||||    |||
Db         82 CAATCTGTGTTGATGCAATCTTATTCTAATATTGAATATATAATAATTGATGGGAAATCT 141

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTAC---------TAGTATGAATATTACA 171
              | |||  |||     |  | | |    || ||          ||||  |  | | |
Db        142 AATGATTCAACACTAGGAATTGTGAAAGAATATATACCTCTATTTAGTGGGCGTTTGAAG 201

Qy        172 CATGTTTATGAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCC 231
              || |    || |||||   || || |||||||| ||||||||||   ||  |||||
Db        202 TATATAAGCGAGAAAGATTCTGGTATCTATGATGCCATGAATAAGGGGATAAAATTGGCT 261

Qy        232 AAAGGCGACTTAATACATTATTTAAACGCCGGCGAT 267
              | || ||  ||||   |  | ||||  | |  |||
Db        262 AGTGGGGATATAATTGGTATTATAAATTCAGATGAT 297
```

RESULT 4
US-11-098-686-8881
; Sequence 8881, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8881
; LENGTH: 966
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-8881

Query Match 7.3%; Score 54.4; DB 17; Length 966;
Best Local Similarity 50.4%; Pred. No. 2.7;
Matches 133; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
Qy          4 AAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGT 63
              | ||||||  |||| |      || |||||||  |    | || |      | ||| |   |
Db         19 ATAATATCCATTATTATACCTACCTATAATATGGAGCCATGGATTACGGACACTATTCAT 78

Qy         64 AGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACA 123
              ||| | | ||     |      |||||||  ||| ||  ||||||  ||||||||
Db         79 AGTGTGTTAAAGCAATCCTTCTCTAATTTTGAGTGTATCATTATTGATGATGGCTCTACT 138

Qy        124 GATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAA 183
              |||   ||   | |   |||   | |  |  |  |  |      | || | |
Db        139 GATAATACTGTACCTACTATTCAACAAATACATGATCCAAGATTGCATATTCTTCAACAG 198

Qy        184 AAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTA 243
              | | || ||||  | | || ||| | |||||  || |  |      || |||||  | |
Db        199 ACAAATCAAGGTGTCTCTGCTGCTAGGAATACAGGTCTTAGTCATGCAAAAGGACAATAT 258
```
```

```
Qy        244 ATACATTATTTAAACGCCGGCGAT 267
              ||    | | |  | |  |  |||
Db        259 ATTGCATTTCTTGATGGAGATGAT 282


RESULT 5
US-11-098-686-8738/c
; Sequence 8738, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
;  APPLICANT: Kapur, Vivek and Gebhart, Connie J.
;  TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
;  TITLE OF INVENTION:  FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
;  FILE REFERENCE: 09531-128001
;  CURRENT APPLICATION NUMBER: US/11/098,686
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: PCT/US03/31318
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: US 60/416,395
;  PRIOR FILING DATE: 2002-10-04
;  NUMBER OF SEQ ID NOS: 11433
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8738
;   LENGTH: 194553
;   TYPE: DNA
;   ORGANISM: Lawsonia intracellularis
US-11-098-686-8738

  Query Match             7.3%;  Score 54.4;  DB 17;  Length 194553;
  Best Local Similarity   50.4%;  Pred. No. 4.7;
  Matches  133;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy          4 AAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGT 63
              | ||||||  |||| |     || |||||||  |    | || |     | ||| |   |
Db     142031 ATAATATCCATTATTATACCTACCTATAATATGGAGCCATGGATTACGGACACTATTCAT 141972

Qy         64 AGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACA 123
              ||| | | ||    |     |       |||||||  ||| ||  ||||||  ||||||||
Db     141971 AGTGTGTTAAAGCAATCCTTCTCTAATTTTGAGTGTATCATTATTGATGATGGCTCTACT 141912

Qy        124 GATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAA 183
              |||   ||    | |   |||     | |   |  |   |   |      | || | |
Db     141911 GATAATACTGTACCTACTATTCAACAAATACATGATCCAAGATTGCATATTCTTCAACAG 141852

Qy        184 AAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTA 243
              | | || ||||  | | || ||| | |||||  || |  |    || |||||  | |
Db     141851 ACAAATCAAGGTGTCTCTGCTGCTAGGAATACAGGTCTTAGTCATGCAAAAGGACAATAT 141792

Qy        244 ATACATTATTTAAACGCCGGCGAT 267
              ||    | | |  | |  |  |||
Db     141791 ATTGCATTTCTTGATGGAGATGAT 141768


RESULT 6
US-11-079-463-2777
; Sequence 2777, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PATH00-03DIV2
;  CURRENT APPLICATION NUMBER: US/11/079,463
;  CURRENT FILING DATE:  2005-03-14
;  PRIOR APPLICATION NUMBER: US 60/128,705
;  PRIOR FILING DATE: 1999-04-09
;  PRIOR APPLICATION NUMBER: US 09/540,209
;  PRIOR FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 2777
;   LENGTH: 783
;   TYPE: DNA
;   ORGANISM: B.fragilis
US-11-079-463-2777

  Query Match             7.0%;  Score 52.4;  DB 18;  Length 783;
  Best Local Similarity   49.3%;  Pred. No. 5.8;
  Matches  137;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy          4 AAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGT 63
              ||| | || || || || ||||| || |||   |   | |  || ||| | ||| |
Db         31 AAATTCTCCGTGATCACAGTGACCTACAATGCCGAAAAGGTATTGGAAGATACTGTGCAA 90

Qy         64 AGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACA 123
              ||  ||          |       | | | | |||  ||| ||  | || || |  || | |
Db         91 AGCGTAATTTCGCAGACCTATCATCATGTAGAGTATATCATCATCGACGGTGCATCGAAA 150

Qy        124 GATGGAACGAATCGTGTCATTAGTAGATTTACTAGTATGAATATTACACATGTTTATGAA 183
              || || ||         |  | |   |||             |  |    |   ||    |||
Db        151 GACGGGACTTTAGAGATTGTCAACCGATACCGCGACCGCATCAACCAATTGGTGAGCGAA 210

Qy        184 AAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGCGACTTA 243
              ||  ||||  | ||||| || ||||| || ||        |||| |  |||||||
Db        211 CCGGACAAAGGACTTTATGACGCAATGAACAAAGGAATCGCCCTGGCAACGGGCGACTAT 270

Qy        244 ATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGA 281
              |    | | | || ||||| || ||  |    | |||
Db        271 CTCTGCTTTCTGAATGCCGGAGACAGTTTCCACGAAGA 308


RESULT 7
US-11-079-463-838
; Sequence 838, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRAC
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PATH00-03DIV2
;  CURRENT APPLICATION NUMBER: US/11/079,463
;  CURRENT FILING DATE:  2005-03-14
;  PRIOR APPLICATION NUMBER: US 60/128,705
;  PRIOR FILING DATE: 1999-04-09
;  PRIOR APPLICATION NUMBER: US 09/540,209
;  PRIOR FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 838
;   LENGTH: 891
;   TYPE: DNA
;   ORGANISM: B.fragilis
US-11-079-463-838

  Query Match             6.6%;  Score 49.4;  DB 18;  Length 891;
  Best Local Similarity   50.2%;  Pred. No. 19;
  Matches  156;  Conservative    0;  Mismatches  146;  Indels    9;  Gaps    1;

Qy          1 ATGAAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTA 60
              ||||  ||| ||||  | ||   ||| | ||||   ||       ||||   | |   |
Db          7 ATGACTATAACAGTGTTTACGCCGACATTTAATCGGGCCACTCTGTTACCTAGATTATAT 66

Qy         61 AGTAGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCT 120
              ||||||   | |  || | |      |||||||    |||| || ||||  ||   |
Db         67 GAAAGTTTAGTTAATCAAACATTTCTTGATTTTGAGTGGCTTATTGTAGATGATGGAAGT 126

Qy        121 ACAGATGGAACGAATCGTGTCATTAGTAGATTTAC---------TAGTATGAATATTACA 171
              || ||||  ||   |    | ||       | |||          ||  ||  ||||   |
Db        127 ACTGATGATACATTTAACTTTATAGAATCAATTAAGGAAAACGATAAGATTGATATACAA 186

Qy        172 CATGTTTATGAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCC 231
              ||   ||| | ||  ||| || | | ||| ||| || |||  |||    | | | ||
Db        187 TATTACTATCAGAATAATGCTGGTAAACATGCTGCTATTAATTGGGGAGTAGAGCTAGCT 246

Qy        232 AAAGGCGACTTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAA 291
              ||||||||  |   |   | || |    ||    | ||||  |   |     |   |
Db        247 AAAGGCGATCTTTTTTTTATTGTAGATAGTGATGAGGTTATGATTGAATCTGGGTTACAG 306
```

```
Qy        292 AATATCAAAGA 302
              | |||   |||
Db        307 ACTATAGTAGA 317


RESULT 8
US-11-045-004-1
; Sequence 1, Application US/11045004
; Publication No. US20060078901A1
; GENERAL INFORMATION:
;  APPLICANT: BUCHRIESER, CARMEN
;  APPLICANT:  FRANGEUL, LIONEL
;  APPLICANT:  COUVE, ELISABETH
;  APPLICANT:  RUSNIOK, CHRISTOPHE
;  APPLICANT:  FSIHI, HAFIDA
;  APPLICANT:  DEHOUX, PIERRE
;  APPLICANT:  DUSSURGET, OLIVIER
;  APPLICANT:  CHETOUANI, FARID
;  APPLICANT:  NEDJARI, HAFED
;  APPLICANT:  GLASER, PHILIPPE
;  APPLICANT:  KUNST, FRANCK
;  APPLICANT:  COSSART, PASCALE
;  APPLICANT:  DANIELS, JUSTIN
;  APPLICANT:  GOEBEI, WERNER
;  APPLICANT:  KREFT, JURGEN
;  APPLICANT:  KUHN, MICHAEL
;  APPLICANT:  NG, EVA
;  APPLICANT:  VAZQUEZ-BOLAND, ANTONIO
;  APPLICANT:  DOMINGUEZ-BERNAL, GUSTAVO
;  APPLICANT:  GARRIDO-GARCIA, PATRICIA
;  APPLICANT:  TIERREZ-MARTINEZ, ALBERTO
;  APPLICANT:  AMEND, ALEXANDRA
;  APPLICANT:  CHAKRABORTY, TRINAD
;  APPLICANT:  DOMANN, EUGEN
;  APPLICANT:  HAIN, THORSTEN
;  APPLICANT:  BERCHE, PATRICK
;  APPLICANT:  CHARBIT, ALAIN
;  APPLICANT:  DURANT, LIONEL
;  APPLICANT:  PEREZ-DIAZ, JOSE-CLAUDIO
;  APPLICANT:  BAQUERO, FERNANDO
;  APPLICANT:  GARCIA DEL PORTILLO, FRANCISCO
;  APPLICANT:  GOMEZ-LOPEZ, NURIA
;  APPLICANT:  MADUENIO, ENCARNA
;  APPLICANT:  PABLOS, BETRIZ DE
;  APPLICANT:  WEHLAND, JURGEN
;  APPLICANT:  KARST, UWE
;  APPLICANT:  ENTIAN, KARL-DIETER
;  APPLICANT:  HAUF, JORG
;  APPLICANT:  ROSE, MATTHIAS
;  APPLICANT:  VOSS, HAMUT
;  TITLE OF INVENTION: LISTERIA MONOCYTOGENES GENOME, POLYPEPTIDES AND USES
;  FILE REFERENCE: 05394.0018-02
;  CURRENT APPLICATION NUMBER: US/11/045,004
;  CURRENT FILING DATE:  2005-01-28
;  PRIOR APPLICATION NUMBER: 10/637,657
;  PRIOR FILING DATE: 2003-08-11
;  PRIOR APPLICATION NUMBER: 10/257,023
;  PRIOR FILING DATE: 2002-10-08
;  PRIOR APPLICATION NUMBER: PCT/FR01/01118
;  PRIOR FILING DATE: 2001-04-11
;  PRIOR APPLICATION NUMBER: FR 00/04,629
;  PRIOR FILING DATE: 2000-04-11
;  NUMBER OF SEQ ID NOS: 2854
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
;   LENGTH: 2944528
;   TYPE: DNA
;   ORGANISM: Listeria monocytogenes
US-11-045-004-1

  Query Match             6.4%;  Score 47.8;  DB 18;  Length 2944528;
  Best Local Similarity   48.7%;  Pred. No. 81;
  Matches  164;  Conservative    0;  Mismatches  167;  Indels    6;  Gaps    1;

Qy          4 AAAATATCAGTTATAACTGTGACTTATAATAATGCTGAAGGGTTAGAAAAAACTTTAAGT 63
              ||| ||   ||  | | ||     || ||| | ||||     |||||  |  ||||
Db    1125488 AAATTAAGTGTGGTCATTGCCGTATACAATGTTTCTGAGTATATAGAAGCATGCTTAACA 1125547

Qy         64 AGTTTATCAATTTTAAAAATAAAACCTTTTGAGATTATTATAGTTGATGGCGGCTCTACA 123
                | | |     |||    ||   | | ||  | |||||  ||  |||  || || ||
Db    1125548 TCAGTCTTAGAACAAAACTACAAGAATGTGGAATTAATTATTGTCAATGATGGTTCAACT 1125607

Qy        124 GATGGAACGAATCGTGTCATTAGTAGATT------TACTAGTATGAATATTACACATGTT 177
              ||||  ||   ||       |  | |             || |||| |||       ||
Db    1125608 GATGATACCTTTCAAAAAGTAGAAAAACTAGTAGACGGACATACGAATGTTATTTTGATT 1125667

Qy        178 TATGAAAAAGATGAAGGGATATATGATGCGATGAATAAGGGCCGAATGTTGGCCAAAGGC 237
               || || || |  ||||  | | || ||     |||||  ||   ||     || | |||
Db    1125668 GATCAAGAAAACCAAGGTGTTTCTGCTGTCCGAAATAATGGTGTAAAAGCCGCTACAGGT 1125727

Qy        238 GACTTAATACATTATTTAAACGCCGGCGATAGCGTAATTGGAGATATATATAAAAATATC 297
              || |   ||  | | | | | | | |||||  | |   |  ||||||  | | |||| ||
Db    1125728 GATTATCTAGGCTTTGTCGATGGAGACGATATCATTGCTAAAGATATGTTTCAAAAAATG 1125787

Qy        298 AAAGAGCCATGTTTGATTAAAGTTGGCCTTTTCGAAA 334
              |  ||  | | |   ||     | |||    | | | ||
Db    1125788 ATGGATTCTTTTGAAATGTCGGATGCAGATATGGTAA 1125824


RESULT 9
US-09-925-065A-674172/c
; Sequence 674172, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 674172
;   LENGTH: 3222
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-674172

  Query Match             6.3%;  Score 47.2;  DB 7;  Length 3222;
  Best Local Similarity   48.8%;  Pred. No. 52;
  Matches  127;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy        417 TTCAGAATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTT 476
              || | | |   ||  ||   | |||   |     |||| ||| |   | | |   | ||
Db       1612 TTTATATTTCTATAAAAATAAATAAAATGGCAAAAATTACAAGTTCTCTGAGTATTAGCT 1553

Qy        477 TCCTGAAGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGG 536
              || | |    | || ||   | | ||   ||   | || || | | |   || |  ||||
Db       1552 TCTTCATCTATAAAAATTGTTCTATTCACTTTACTTGGTTTTTCTGAGTAATTAACTGGG 1493

Qy        537 GGGAGTATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGA 596
                ||||  | ||| ||   ||| ||   | ||  | ||  |   |  | | ||| | |
Db       1492 TAATGTATGATGAAATAATTAAATATTTTTAAATTTAAAGTGTTATATTAATTTGTATAA 1433

Qy        597 AAAAAATAAAAAAAACCTTATTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCC 656
```

```
Db       1432 ACATATGTACACAAAGAAAATGTATTTCTCTATAAACATAGAAATACAAATTTAAAGACT 1373

Qy        657 TGAACGTTTAAGAAGAGTAT 676
Db       1372 AGAATTTTTCCTTAGAGTCT 1353


RESULT 10
US-10-301-480-568794
; Sequence 568794, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 568794
;   LENGTH: 990
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-568794

  Query Match             6.2%;  Score 46.4;  DB 12;  Length 990;
  Best Local Similarity   46.6%;  Pred. No. 62;
  Matches  181;  Conservative    0;  Mismatches  206;  Indels    1;  Gaps    1;

Qy        266 ATAGCGTAATTGGAGATATATATAAAAATATCAAAGAGCCATGTTTGATTAAAGTTGGCC 325
Db        311 ATATAATATTATATTATATATTATATTAAATATAACATAATAGTTTGTATTAAATATATT 370

Qy        326 TTTTCGAAAATGATAAACTTCTGGGATTTTCTTCTATAACCCATTCAAATACAGGGTATT 385
Db        371 ATATATAATATAATATAATAATGTTATATAATGTAATAATATATTATTTTATAATGTAAC 430

Qy        386 GTCATCAAGGGGTGATTTTCCCAAAGAATCATTCAGAATATGATCTAAGGTATAAAATAT 445
Db        431 ATAATTATATTATTTTATAATATATAATTTATTTAAAATATAATATATAATATAATGTAA 490

Qy        446 GTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGA 505
Db        491 TATATGATTATATATTGTATTATGTTATATGTAATATATTATATATAA-TATATGATATA 549

Qy        506 TTACTTCGGGTTATGTAAAATATGATATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAA 565
Db        550 TTAATAATATATATAAATAATTTATATATTATATATAATTTATATATATTATATTTTAAA 609

Qy        566 GAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTA 625
Db        610 ATGTAATTTATATATATAAATTATATATTATAAAATATAAAATATATAGTATATAATATA 669

Qy        626 TTCCAATTTCAATAATCAAAATTTTATT 653
Db        670 TAATACATTATTTTGTTCATTTTAAATT 697


RESULT 11
US-10-301-480-1182203
; Sequence 1182203, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1182203
;   LENGTH: 990
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1182203

  Query Match             6.2%;  Score 46.4;  DB 12;  Length 990;
  Best Local Similarity   46.6%;  Pred. No. 62;
  Matches  181;  Conservative    0;  Mismatches  206;  Indels    1;  Gaps    1;

Qy        266 ATAGCGTAATTGGAGATATATATAAAAATATCAAAGAGCCATGTTTGATTAAAGTTGGCC 325
Db        311 ATATAATATTATATTATATATTATATTAAATATAACATAATAGTTTGTATTAAATATATT 370

Qy        326 TTTTCGAAAATGATAAACTTCTGGGATTTTCTTCTATAACCCATTCAAATACAGGGTATT 385
Db        371 ATATATAATATAATATAATAATGTTATATAATGTAATAATATATTATTTTATAATGTAAC 430

Qy        386 GTCATCAAGGGGTGATTTTCCCAAAGAATCATTCAGAATATGATCTAAGGTATAAAATAT 445
Db        431 ATAATTATATTATTTTATAATATATAATTTATTTAAAATATAATATATAATATAATGTAA 490

Qy        446 GTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGA 505
Db        491 TATATGATTATATATTGTATTATGTTATATGTAATATATTATATATAA-TATATGATATA 549

Qy        506 TTACTTCGGGTTATGTAAAATATGATATGGGGGGAGTATCTTCAAAAAAAAGAATTTTAA 565
Db        550 TTAATAATATATATAAATAATTTATATATTATATATAATTTATATATATTATATTTTAAA 609

Qy        566 GAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAACCTTATTAAGTTTA 625
Db        610 ATGTAATTTATATATATAAATTATATATTATAAAATATAAAATATATAGTATATAATATA 669

Qy        626 TTCCAATTTCAATAATCAAAATTTTATT 653
Db        670 TAATACATTATTTTGTTCATTTTAAATT 697


RESULT 12
US-10-301-480-293547
; Sequence 293547, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 293547
;   LENGTH: 579
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-293547

  Query Match             6.2%;  Score 46;  DB 12;  Length 579;
  Best Local Similarity   50.5%;  Pred. No. 69;
  Matches  109;  Conservative    1;  Mismatches  106;  Indels    0;  Gaps    0;
```

```
Qy        423 ATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGA 482
              ||   ||||| | || |  ||| || ||||  | |   |     |||  |    | :|
Db        283 ATGAAATCTATGATACAGTATACATGGTGATCTTCATGCTTAATGAGACATACCACAYGC 342

Qy        483 AGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGAGT 542
              | | |   ||| | |       |   |||  || |  ||||  | ||||  |    |
Db        343 AAGCTGCTGATATATTGGCCAGTGATCTTTTTGTAAGATAAATAAAGATAAAGAATCATC 402

Qy        543 ATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAA 602
                ||  |   ||  |||||||   ||| |     ||     | ||| ||| ||  | |
Db        403 TATTTTCATTTTAATTATTTTAAATAATATATTAGGTGTATTTAATATATTTTAATTATA 462

Qy        603 TAAAAAAAACCTTATTAAGTTTATTCCAATTTCAAT 638
              || |||| |  || || | || |||  ||||| |||
Db        463 TACAAAATAATTTGTTTATTTAATTAAAATTTTAAT 498
```

RESULT 13
US-10-301-480-906956
; Sequence 906956, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 906956
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-906956

Query Match 6.2%; Score 46; DB 12; Length 579;
Best Local Similarity 50.5%; Pred. No. 69;
Matches 109; Conservative 1; Mismatches 106; Indels 0; Gaps 0;

```
Qy        423 ATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGA 482
              ||   ||||| | || |  ||| || ||||  | |   |     |||  |    | :|
Db        283 ATGAAATCTATGATACAGTATACATGGTGATCTTCATGCTTAATGAGACATACCACAYGC 342

Qy        483 AGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGAGT 542
              | | |   ||| | |       |   |||  || |  ||||  | ||||  |    |
Db        343 AAGCTGCTGATATATTGGCCAGTGATCTTTTTGTAAGATAAATAAAGATAAAGAATCATC 402

Qy        543 ATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAA 602
                ||  |   ||  |||||||   ||| |     ||     | ||| ||| ||  | |
Db        403 TATTTTCATTTTAATTATTTTAAATAATATATTAGGTGTATTTAATATATTTTAATTATA 462

Qy        603 TAAAAAAAACCTTATTAAGTTTATTCCAATTTCAAT 638
              || |||| |  || || | || |||  ||||| |||
Db        463 TACAAAATAATTTGTTTATTTAATTAAAATTTTAAT 498
```

RESULT 14
US-09-925-065A-205659
; Sequence 205659, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 205659
; LENGTH: 587
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-205659

Query Match 6.2%; Score 46; DB 7; Length 587;
Best Local Similarity 50.5%; Pred. No. 69;
Matches 109; Conservative 1; Mismatches 106; Indels 0; Gaps 0;

```
Qy        423 ATATGATCTAAGGTATAAAATATGTGCTGATTATAAGCTTATTCAAGAGGTGTTTCCTGA 482
              ||   ||||| | || |  ||| || ||||  | |   |     |||  |    | :|
Db        283 ATGAAATCTATGATACAGTATACATGGTGATCTTCATGCTTAATGAGACATACCACAYGC 342

Qy        483 AGGGTTAAGATCTCTATCTTTGATTACTTCGGGTTATGTAAAATATGATATGGGGGGAGT 542
              | | |   ||| | |       |   |||  || |  ||||  | ||||  |    |
Db        343 AAGCTGCTGATATATTGGCCAGTGATCTTTTTGTAAGATAAATAAAGATAAAGAATCATC 402

Qy        543 ATCTTCAAAAAAAAGAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAA 602
                ||  |   ||  |||||||   ||| |     ||     | ||| ||| ||  | |
Db        403 TATTTTCATTTTAATTATTTTAAATAATATATTAGGTGTATTTAATATATTTTAATTATA 462

Qy        603 TAAAAAAAACCTTATTAAGTTTATTCCAATTTCAAT 638
              || |||| |  || || | || |||  ||||| |||
Db        463 TACAAAATAATTTGTTTATTTAATTAAAATTTTAAT 498
```

RESULT 15
US-11-117-187-211
; Sequence 211, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE
; APPLICANT: COPENHAVER, GREGORY
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD:309US
; CURRENT APPLICATION NUMBER: US/11/117,187
; CURRENT FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/09/531,120
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 211
; LENGTH: 1082144
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-11-117-187-211

Query Match 6.2%; Score 46; DB 17; Length 1082144;
Best Local Similarity 52.6%; Pred. No. 1.5e+02;
Matches 100; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```
Qy        557 GAATTTTAAGAGATAAAGAGCTTGCCAAAATTATGTTTGAAAAAAATAAAAAAAACCTTA 616
              ||| |   ||| |||| | |     | |||||| ||     ||||    || ||  | |
Db     149742 GAAATCCTAGATATAACAATCAAAATATAATTATATTATCTTAAAAACTTAACAAAATAA 149801

Qy        617 TTAAGTTTATTCCAATTTCAATAATCAAAATTTTATTCCCTGAACGTTTAAGAAGAGTAT 676
              ||   ||| |  |||    ||||| |||||||  | |   | ||   | |  ||| | |
Db     149802 ATATTATTAATAAAATAAACATAATTAAAATTTAAATTATTTAATAATAATCTAGATTTT 149861
```

```
Qy        677 TGCGGAAAATGCAATATATTTGTCTAACTTTATTCTTCATGAAGAATAGTTCACCATATG 736
              |  ||    |  || |||  | |  || |||| || |  |  || | |  | ||||
Db     149862 AACTTAATGAACTGTACATTATTGTTCCTATATTTTTAAACAGTAAAAATCAAACATAGT 149921

Qy        737 ATAATGAATA 746
              |  |   |||
Db     149922 AAGAGTTATA 149931
```

Search completed: May 11, 2006, 08:30:19
Job time : 393.013 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:45:22 ; Search time 4018.18 Seconds
(without alignments)
15660.244 Million cell updates/sec

Title: US-09-423-093B-2_COPY_5365_6471
Perfect score: 1107
Sequence: 1 attcttttaccgagaaatga.....aaattgaatcccgatcttaa 1107

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1107	100.0	14002	1	AF061251	AF061251 Escherici
2		1107	100.0	14024	6	BD085555	BD085555 Nucleic a
3		1107	100.0	14029	6	AX029262	AX029262 Sequence
4	c	1107	100.0	14181	6	BD184789	BD184789 Nucleic a
5	c	1107	100.0	14187	6	AR204225	AR204225 Sequence
6	c	1107	100.0	14187	6	AR637574	AR637574 Sequence
7		1107	100.0	31482	1	AB008676	AB008676 Escherich
8	c	1107	100.0	110000	1	AE005174_28	Continuation (29 o
9	c	1107	100.0	110000	1	BA000007_27	Continuation (28 o
10		573.2	51.8	14323	1	AY730593	AY730593 Salmonell
11		476.6	43.1	14224	1	AY730592	AY730592 Citrobact
12	c	76.6	6.9	155204	14	AC007926	AC007926 Trypanoso
13		75.6	6.8	225581	14	BX537105	BX537105 Danio rer
14		73.6	6.6	115990	5	BX890614	BX890614 Zebrafish
15		73.6	6.6	155885	14	CR848672	CR848672 Danio rer
16	c	72	6.5	110000	14	CT009752_3	Continuation (4 of
17		72	6.5	144604	14	BX899179	BX899179 Danio rer
18		72	6.5	145553	5	AL935272	AL935272 Zebrafish
19		71.6	6.5	79882	14	BX908758_5	Continuation (6 of
20	c	71.6	6.5	191146	5	BX530070	BX530070 Zebrafish
21		71	6.4	2426	15	SDU49822	U49822 Saccharomyc
22		70.2	6.3	154561	5	CR391989	CR391989 Zebrafish
23	c	69.8	6.3	144375	14	CR354586	CR354586 Danio rer
24	c	69.2	6.3	163843	5	BX510939	BX510939 Zebrafish
25	c	68.6	6.2	110000	14	CT009752_6	Continuation (7 of
26	c	68.6	6.2	129240	8	AC084128	AC084128 Homo sapi
27		68.2	6.2	142168	14	CR388191	CR388191 Danio rer
28	c	67.6	6.1	151341	5	AL929536	AL929536 Zebrafish
29		67.2	6.1	56494	2	AE003846_3	Continuation (4 of
30		66.6	6.0	110000	14	CT009752_4	Continuation (5 of
31		66.6	6.0	110000	14	CT009752_5	Continuation (6 of
32		65.8	5.9	173019	5	BX321884	BX321884 Zebrafish
33	c	65.4	5.9	182464	5	BX569779	BX569779 Zebrafish
34	c	65	5.9	110000	14	AL954295_2	Continuation (3 of
35	c	65	5.9	125031	5	BX248494	BX248494 Zebrafish
36	c	64.8	5.9	4865	2	AF364131	AF364131 Anopheles
37		64.8	5.9	64707	2	AC115607	AC115607 Dictyoste
38		64.6	5.8	5164	2	AF482393	AF482393 Dictyoste
39	c	64.6	5.8	152811	8	AC006020	AC006020 Homo sapi
40		64.4	5.8	177575	5	AL953867	AL953867 Zebrafish
41		63.6	5.7	139980	14	CR855330	CR855330 Danio rer
42		63.6	5.7	182464	5	BX569779	BX569779 Zebrafish
43		63	5.7	34340	14	BX957360	BX957360 Danio rer
44		62.8	5.7	110000	14	CT009753_07	Continuation (8 of
45		62.6	5.7	160256	5	BX950187	BX950187 Zebrafish

ALIGNMENTS

RESULT 1
AF061251
LOCUS AF061251 14002 bp DNA linear BCT 20-AUG-1998
DEFINITION Escherichia coli serotype O157:H7 O antigen gene cluster.
ACCESSION AF061251
VERSION AF061251.1 GI:3435170
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 14002)
AUTHORS Wang,L. and Reeves,P.R.
TITLE Organisation of Escherichia coli O157 O antigen gene cluster and identification of its specific genes
JOURNAL Infect. Immun. 66 (8), 3545-3551 (1998)
PUBMED 9673232
REFERENCE 2 (bases 1 to 14002)
AUTHORS Wang,L.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1998) Microbiology, The University of Sydney, NSW 2006, Australia
FEATURES Location/Qualifiers
source 1. .14002
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="C664-1992"
/serotype="O157:H7"
/db_xref="taxon:562"
gene 60. .842
/gene="wbdN"
CDS 60. .842
/gene="wbdN"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdN"
/protein_id="AAC32339.1"
/db_xref="GI:3435171"
/translation="MNKETVSIIMPVYNGAKTIISSVESIIHQSYQDFVLYIIDDCST
DDTFSLINSRYKNNQKIRILRNKTNLGVAESRNYGIEMATGKYISFCDADDLWHEKKL
ERQIEVLNNECVDVVCSNYYVIDNNRNIVGEVNAPHVINYRKMLMKNYIGNLTGIYNA

NKLGKFYQKKIGHEDYLMWLEIINKTNGAICIQDNLAYYMRSNNSLSGNKIKAAKWTW
SIYREHLHLSFPKTLYYFLLYASNGVMKKITHSLLRRKETKK"
gene 839. .2023
/gene="wzy"
CDS 839. .2023
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O antigen polymerase Wzy"
/protein_id="AAC32340.1"
/db_xref="GI:3435172"
/translation="MKSAAKLIFLFLFTLYSLQLYGVIIDDRITNFDTKVLTSIIIIF
QIFFVLLFYLTIINERKQQKKFIVNWELKLILVFLFVTIEIAAVVLFLKEGIPIFDDD
PGGAKLRIAEGNGLYIRYIKYFGNIVVFALIILYDEHKFKQRTIIFVYFTTIALFGYR
SELVLLILQYILITNILSKDNRNPKIKRIIGYFLLVGVVCSLFYLSLGQDGEQNDSYN
NMLRIINRLTIEQVEGVPYVVSESIKNDFFPTPELEKELKAIINRIQGIKHQDLFYGE
RLHKQVFGDMGANFLSVTTYGAELLVFFGFLCVFIIPLGIYIPFYLLKRMKKTHSSIN
CAFYSYIIMILLQYLVAGNASAFFFGPFLSVLIMCTPLILLHDTLKRLSRNENISYNC
DL"
gene 1992. .2738
/gene="wbdO"
CDS 1992. .2738
/gene="wbdO"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdO"
/protein_id="AAC32341.1"
/db_xref="GI:3435173"
/translation="MKISVITVTYNNAEGLEKTLSSLSILKIKPFEIIIVDGGSTDGT
NRVISRFTSMNITHVYEKDEGIYDAMNKGRMLAKGDLIHYLNAGDSVIGDIYKNIKEP
CLIKVGLFENDKLLGFSSITHSNTGYCHQGVIFPKNHSEYDLRYKICADYKLIQEVFP
EGLRSLSLITSGYVKYDMGGVSSKKRILRDKELAKIMFEKNKKNLIKFIPISIIKILF
PERLRRVLRKMQYICLTLFFMKNSSPYDNE"
gene 2725. .4116
/gene="wzx"
CDS 2725. .4116
/gene="wzx"
/codon_start=1
/transl_table=11
/product="O antigen flippase Wzx"
/protein_id="AAC32342.1"
/db_xref="GI:3435174"
/translation="MIMNKIKKILKFCTLKKYDTSSALGREQERYRIISLSVISSLIS
KILSLLSLILTVSLTLPYLGQERFGVWMTITSLGAALTFLDLGIGNALTNRIAHSFAC
GKNLKMSRQISGGLTLLAGLSFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKV
FVIIFGLGIYSNGVQKVYMGIQKAYISNIVNAIFILLSIITLVISSKLHAGLPVLIVS
TLGIQYISGIYLTINLIIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSG
DNFIISITLGVTYVAVFSITQRLFQISTVPLTIYNIPLWAAYADAHARNDTQFIKKTL
RTSLKIVGISSFLLAFILVVFGSEVVNIWTEGKIQVPRTFIIAYALWSVIDAFSNTFA
SFLNGLNIVKQQMLAVVTLILIAIPAKYIIVSHFGLTVMLYCFIFIYIVNYFIWYKCS
FKKHIDRQLNIRG"
gene 4113. .5213
/gene="per"
CDS 4113. .5213
/gene="per"
/codon_start=1
/transl_table=11
/product="perosamine synthetase Per"
/protein_id="AAC32343.1"
/db_xref="GI:3435175"
/translation="MKMKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKF
AEQNHVQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFV
DSDNETWQMSVSDIEQKITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEA
FGSKYKGKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHR
QYWHDVIGYNYRMTNICAAIGLAQLEQADDFISRKREIADIYKKNINSLVQVHKESKD
VFHTYWMVSILTRTAEEREELRNHLADKLIETRPVFYPVHTMPMYSEKYQKHPIAEDL
GWRGINLPSFPSLSNEQVIYICESINEFYSDK"
gene 5238. .6452
/gene="wbdP"
CDS 5238. .6452
/gene="wbdP"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdP"
/protein_id="AAC32344.1"
/db_xref="GI:3435176"
/translation="MKIALNSDGFYEWGGGIDFIKYILSILETKPEICIDILLPRNDI
HSLIREKAFPFKSILKAILKRERPRWISLNRFNEQYYRDAFTQNNIETNLTFIKSKSS
AFYSYFDSSDCDVILPCMRVPSGNLNKKAWIGYIYDFQHCYYPSFFSKREIDQRNVFF
KLMLNCANNIIVNAHSVITDANKYVGNYSAKLHSLPFSPCPQLKWFADYSGNIAKYNI
DKDYFIICNQFWKHKDHATAFRAFKIYTEYNPDVYLVCTGATQDYRFPGYFNELMVLA
KKLGIESKIKILGHIPKLEQIELIKNCIAVIQPTLFEGGPGGGVTFDAIALGKKVILS
DIDVNKEVNCGDVYFFQAKNHYSLNDAMVKADESKIFYEPTTLIELGLKRRNACADFL
LDVVKQEIESRS"
gene 6472. .7590
/gene="gmd"
CDS 6472. .7590
/gene="gmd"
/codon_start=1
/transl_table=11
/product="GDP-D-mannose dehydratase Gmd"
/protein_id="AAC32345.1"
/db_xref="GI:3435177"
/translation="MTKVALITGVTGQDGSYLAEFLLDKGYEVHGIKRRASSFNTERI
DHIYQDPHGSNPNFHLHYGDLTDSSNLTRILKEVQPDEVYNLAAMSHVAVSFESPEYT
ADVDAIGTLRLLEAIRFLGLENKTRFYQASTSELYGLVQEIPQKESTPFYPRSPYAVA
KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRGLANIAQGLESCLYL
GNMDSLRDWGHAKDYVRMQWLMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLGIKMSF
VGKGIEEKGIVDSVEGQDAPGVKPGDVIVAVDPRYFRPAEVDTLLGDPSKANLKLGWR
PEITLAEMISEMVAKDLEAAKKHSLLKSHGFSVSLALE"
gene 7587. .8558
/gene="fcl"
CDS 7587. .8558
/gene="fcl"
/codon_start=1
/transl_table=11
/product="fucose synthetase Fcl"
/protein_id="AAC32346.1"
/db_xref="GI:3435178"
/translation="MMMNKQRIFIAGHQGMVGSAITRRLKQRDDVELVLRTRDELNLL
DSSAVLDFFSSQKIDQVYLAAAKVGGILANSSYPADFIYENIMIEANVIHAAHKNNVN
KLLFLGSSCIYPKLAHQPIMEDELLQGKLEPTNEPYAIAKIAGIKLCESYNRQFGRDY
RSVMPTNLYGPNDNFHPSNSHVIPALLRRFHDAVENNSPNVVVWGSGTPKREFLHVDD
MASASIYVMEMPYDIWQKNTKVMLSHINIGTGIDCTICELAETIAKVVGYKGHITFDT
TKPDGAPRKLLDVTLLHQLGWNHKITLHKGLENTYNWFLENQLQYRG"
gene 8561. .9070
/gene="wbdQ"
CDS 8561. .9070
/gene="wbdQ"
/codon_start=1
/transl_table=11
/product="WbdQ"
/protein_id="AAC32347.1"
/db_xref="GI:3435179"
/translation="MFLHSQDFATIVRSTPLISIDLIVENEFGEILLGKRINRPAQGY
WFVPGGRVLKDEKLQTAFERLTEIELGIRLPLSVGKFYGIWQHFYEDNSMGGDFSTHY
IVIAFLLKLQPNILKLPKSQHNAYCWLSRAKLINDDDVHYNCRAYFNNKTNDAIGLDN
KDIICLMRQ"
gene 9052. .10500
/gene="manC"
CDS 9052. .10500
/gene="manC"
/codon_start=1
/transl_table=11
/product="mannose-1-P guanosyltransferase ManC"
/protein_id="AAC32348.1"
/db_xref="GI:3435180"
/translation="MSDAPIIAVVMAGGTGSRLWPLSRELYPKQFLQLSGDNTLLQTT
LLRLSGLSCQKPLVITNEQHRFVVAEQLREINKLNGNIILEPCGRNTAPAIAISAFHA
LKRNPQEDPLLLVLAADHVIAKESVFCDAIKNATPIANQGKIVTFGIIPEYAETGYGY
IERGELSVPLQGHENTGFYYVNKFVEKPNRETAELYMTSGNHYWNSGIFMFKASVYLE
ELRKFRPDIYNVCEQVASSSYIDLDFIRLSKEQFQDCPAESIDFAVMEKTEKCVVCPV
DIGWSDVGSWQSLWDISLKSKTGDVCKGDILTYDTKNNYIYSESALVAAIGIEDMVIV
QTKDAVLVSKKSDVQHVKKIVEMLKLQQRTEYISHREVFRPWGKFDSIDQGERYKVKK
IIVKPGEGLSLRMHHHRSEHWIVLSGTAKVTLGDKTKLVTANESIYIPLGAAYSLENP
GIIPLNLIEVSSGDYLGEDDIIRQKERYKHED"

```
     gene            10504. .11874
                     /gene="manB"
     CDS             10504. .11874
                     /gene="manB"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphomannomutase ManB"
                     /protein_id="AAC32349.1"
                     /db_xref="GI:3435181"
                     /translation="MKSLTCFKAYDIRGKLGEELNEDIAWRIGRAYGEFLKPKTIVLG
                     GDVRLTSEALKLALAKGLQDAGVDVLDIGMSGTEEIYFATFHLGVDGGIEVTASHNPM
                     DYNGMKLVREGARPISGDTGLRDVQRLAEANDFPPVDETKRGRYQQINLRDAYVDHLF
                     GYINVKNLTPLKLVINSGNGAAGPVVDAIEARFKALGAPVELIKVHNTPDGNFPNGIP
                     NPLLPECRDDTRNAVIKHGADMGIAFDGDFDRCFLFDEKGQFIEGYYIVGLLAEAFLE
                     KNPGAKIIHDPRLSWNTVDVVTAAGGTPVMSKTGHAFIKERMRKEDAIYGGEMSAHHY

  Query Match           100.0%;  Score 1107;  DB 1;  Length 14002;
  Best Local Similarity  100.0%;  Pred. No. 4.5e-168;
  Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5346 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 5405

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5406 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 5465

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5466 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 5525

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5526 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 5585

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5586 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 5645

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5646 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 5705

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5706 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 5765

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5766 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 5825

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5826 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 5885

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5886 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 5945

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5946 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 6005

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6006 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 6065

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6066 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 6125

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6126 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 6185

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6186 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 6245

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6246 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 6305

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6306 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 6365

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6366 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 6425

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
              ||||||||||||||||||||||||||
Db       6426 AAACAAGAAATTGAATCCCGATCTTAA 6452


RESULT 2
BD085555
LOCUS       BD085555               14024 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085555
VERSION     BD085555.1  GI:22631165
KEYWORDS    JP 2001522248-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 14024)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 2 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Escherichia coli
            PN   JP 2001522248-A/2
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14024
            FT                   /organism='Escherichia coli'.
FEATURES             Location/Qualifiers
     source          1. .14024
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match           100.0%;  Score 1107;  DB 6;  Length 14024;
  Best Local Similarity  100.0%;  Pred. No. 4.5e-168;
  Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5365 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 5424

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5425 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 5484

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       5485 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 5544
Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
Db       5545 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 5604
Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
Db       5605 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 5664
Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
Db       5665 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 5724
Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
Db       5725 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 5784
Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
Db       5785 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 5844
Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
Db       5845 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 5904
Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
Db       5905 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 5964
Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
Db       5965 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 6024
Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
Db       6025 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 6084
Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
Db       6085 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 6144
Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
Db       6145 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 6204
Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
Db       6205 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 6264
Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
Db       6265 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 6324
Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
Db       6325 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 6384
Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
Db       6385 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 6444
Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
Db       6445 AAACAAGAAATTGAATCCCGATCTTAA 6471

RESULT 3
AX029262
LOCUS       AX029262               14029 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 2 from Patent WO9850531.
ACCESSION   AX029262
VERSION     AX029262.1  GI:10190113
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 2 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .14029
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
ORIGIN

  Query Match           100.0%;  Score 1107;  DB 6;  Length 14029;
  Best Local Similarity  100.0%;  Pred. No. 4.5e-168;
  Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
Db       5365 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 5424
Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
Db       5425 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 5484
Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
Db       5485 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 5544
Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
Db       5545 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 5604
Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
Db       5605 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 5664
Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
Db       5665 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 5724
Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
Db       5725 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 5784
Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
Db       5785 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 5844
Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
Db       5845 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 5904
Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
Db       5905 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 5964
Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
Db       5965 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 6024
Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
Db       6025 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 6084
Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
Db       6085 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 6144
```

```
Qy       781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6145 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 6204

Qy       841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6205 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 6264

Qy       901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6265 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 6324

Qy       961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6325 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 6384

Qy      1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6385 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 6444

Qy      1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
            ||||||||||||||||||||||||||||
Db      6445 AAACAAGAAATTGAATCCCGATCTTAA 6471
```

```
RESULT 4
BD184789/c
LOCUS       BD184789               14181 bp    DNA     linear   PAT 17-JUN-2003
DEFINITION  Nucleic acid molecule and polypeptide specific to intestinal
            hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
            thereof.
ACCESSION   BD184789
VERSION     BD184789.1  GI:31876989
KEYWORDS    JP 2002355074-A/25.
SOURCE      Escherichia coli O157:H7
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 14181)
  AUTHORS   Hayashi,H., Shinagawa,H., Makino,K., Hayashi,T., Onishi,S.,
            Hattori,M. and Kurokawa,K.
  TITLE     Nucleic acid molecule and polypeptide specific to intestinal
            hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
  JOURNAL   Patent: JP 2002355074-A 25 10-DEC-2002;
            PRESIDENT OF UNIVERSITY OF TSUKUBA
COMMENT     OS   Escherichia coli O157:H7
            PN   JP 2002355074-A/25
            PD   10-DEC-2002
            PF   24-JAN-2002 JP 2002015959
            PI   HIDEO HAYASHI,HIDEO SHINAGAWA,KOZO MAKINO,TETSUYA HAYASHI,SHIN
            PI   ONISHI,
            PI   MASAHIRA HATTORI,KEN KUROKAWA
            PC   C12N15/09,C12N15/09,A61K31/7088,A61K39/00,A61K48/00,A61P31/04,
            PC   C07K14/245,
            PC   C07K16/12,C12M1/00,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/
            PC   02,C12Q1/68,
            PC   G01N33/15,G01N33/50,G01N33/53,G01N33/53,G01N33/566,G01N37/00,
            PC   C12N15/00,
            PC   C12N15/00,C12N5/00
            CC   Nucleic acid molecule and polypeptide specific to intestinal
            CC           hemorrhagic
            CC   pathogenic Escherichia coli O157:H7, and method of use thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14181
            FT                   /organism='Escherichia coli O157:H7'.
FEATURES             Location/Qualifiers
     source          1. .14181
                     /organism="Escherichia coli O157:H7"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:83334"
ORIGIN
```

```
 Query Match             100.0%;  Score 1107;  DB 6;  Length 14181;
 Best Local Similarity   100.0%;  Pred. No. 4.5e-168;
 Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8631 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 8572

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8571 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 8512

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8511 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 8452

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8451 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 8392

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8391 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 8332

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8331 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 8272

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8271 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 8212

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8211 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 8152

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8151 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 8092

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8091 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 8032

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8031 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 7972

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7971 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 7912

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7911 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 7852

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7851 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 7792

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7791 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 7732

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7731 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 7672

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7671 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 7612
```

```
Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7611 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 7552

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
             ||||||||||||||||||||||||||||
Db       7551 AAACAAGAAATTGAATCCCGATCTTAA 7525


RESULT 5
AR204225/c
LOCUS       AR204225               14187 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 121 from patent US 6365723.
ACCESSION   AR204225
VERSION     AR204225.1  GI:21500820
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14187)
  AUTHORS   Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
  TITLE     Sequences of E. coli O157
  JOURNAL   Patent: US 6365723-A 121 02-APR-2002;
FEATURES             Location/Qualifiers
     source          1. .14187
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           100.0%;  Score 1107;  DB 6;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 4.5e-168;
  Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8636 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 8577

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8576 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 8517

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8516 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 8457

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8456 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 8397

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8396 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 8337

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8336 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 8277

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8276 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 8217

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8216 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 8157

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8156 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 8097

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8096 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 8037

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8036 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 7977

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7976 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 7917

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7916 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 7857

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7856 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 7797

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7796 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 7737

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7736 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 7677

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7676 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 7617

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7616 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 7557

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
              ||||||||||||||||||||||||||||
Db       7556 AAACAAGAAATTGAATCCCGATCTTAA 7530


RESULT 6
AR637574/c
LOCUS       AR637574               14187 bp    DNA     linear   PAT 20-APR-2005
DEFINITION  Sequence 121 from patent US 6855814.
ACCESSION   AR637574
VERSION     AR637574.1  GI:62771316
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14187)
  AUTHORS   Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
  TITLE     Sequences of E. coli O157
  JOURNAL   Patent: US 6855814-A 121 15-FEB-2005;
            Wisconsin Alumni Research Foundation; Madison, WI
FEATURES             Location/Qualifiers
     source          1. .14187
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match           100.0%;  Score 1107;  DB 6;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 4.5e-168;
  Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8636 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 8577

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8576 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 8517

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8516 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 8457
```

```
Qy       181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8456 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 8397

Qy       241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8396 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 8337

Qy       301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8336 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 8277

Qy       361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8276 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 8217

Qy       421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8216 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 8157

Qy       481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8156 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 8097

Qy       541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8096 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 8037

Qy       601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8036 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 7977

Qy       661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7976 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 7917

Qy       721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7916 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 7857

Qy       781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7856 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 7797

Qy       841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7796 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 7737

Qy       901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7736 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 7677

Qy       961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7676 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 7617

Qy      1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7616 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 7557

Qy      1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
            ||||||||||||||||||||||||||
Db      7556 AAACAAGAAATTGAATCCCGATCTTAA 7530
```

```
RESULT 7
AB008676
LOCUS       AB008676               31482 bp    DNA     linear   BCT 19-MAY-1999
DEFINITION  Escherichia coli 0157 DNA, map position at 46 min., complete cds.
ACCESSION   AB008676
VERSION     AB008676.1  GI:4867913
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Shimizu,T., Yamasaki,S., Tsukamoto,T. and Takeda,Y.
  TITLE     Analysis of the genes responsible for the O-antigen synthesis in
            enterohaemorrhagic Escherichia coli O157
  JOURNAL   Microb. Pathog. 26 (5), 235-247 (1999)
   PUBMED   10222209
REFERENCE   2  (bases 1 to 31482)
  AUTHORS   Shimizu,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1997) Takeshi Shimizu, International Medical
            Center of Japan, Department of Appropriate Technology Development,
            Research Institute; 1-21-1 Toyama, Shinjuku-ku, Tokyo 162, Japan
            (E-mail:kshimizu@imcj.go.jp, Tel:+81-3-3202-7181,
            Fax:+81-3-3202-7364)
FEATURES             Location/Qualifiers
     source          1. .31482
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="184"
                     /serotype="O157"
                     /db_xref="taxon:562"
                     /map="46 min."
                     /clone="1-4-1"
     gene            505. .1398
                     /gene="galF"
     CDS             505. .1398
                     /gene="galF"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="UTP-glucose-1-phosphate uridylyltransferase"
                     /protein_id="BAA77723.1"
                     /db_xref="GI:4867914"
                     /translation="MTNLKAVIPVAGLGMHMLPATKAIPKEMLPIVDKPMIQYIVDEI
                     VAAGIKEILLVTHASKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNV
                     RQGEPLGLGHSILCARPAIGDNPFVVVLPDVVIDDASADPLRYNLAAMIARFNETGRS
                     QVLAKRMPGDLSEYSVIQTKEPLDREGKVSRIVEFIEKPDQPQTLDSDIMAVGRYVLS
                     ADIWPELERTQPGAWGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKY
                     GLRNLKEGAKFRKGIEKLLSE"
     gene            1717. .2499
                     /gene="wbhA"
     CDS             1717. .2499
                     /gene="wbhA"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyl transferase"
                     /protein_id="BAA77724.1"
                     /db_xref="GI:4867915"
                     /translation="MNKETVSIIMPVYNGAKTIISSVESIIHQSYQDFVLYIIDDCST
                     DDTFSLINSRYKNNQKIRILRNKTNLGVAESRNYGIEMATGKYISFCDADDLWHEKKL
                     ERQIEVLNNECVDVVCSNYYVIDNKRNIVGEVNAPHVINYRKMLMKNYIGNLTGIYNA
                     NKLGKFYQKKIGHEDYLMWLEIINKTNGAICIQDNLAYYMRSNNSLSGNKIKAAKWTW
                     SIYREHLHLSFPKTLYYFLLYASNGVMKKITHSLLRRKETKK"
     gene            3153. .3680
                     /gene="wbhB"
     CDS             3153. .3680
                     /gene="wbhB"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase"
                     /protein_id="BAA77725.1"
                     /db_xref="GI:4867916"
                     /translation="MLRIINRLTIEQVESVPYVVSESIKNDFFPTPELEKELKAIINR
                     IQGIKHQDLFYGERLHKQVFGDMGANFLSVTTYGAELLVFFGFLCVFIIPLGIYIPFY
                     LLKRMKKTHSSINCAFYSYIIMILLQYLVAGNASAFFFGPFLSVLIMCTPLILLHDTL
                     KRLSRNENISYNCDL"
     gene            3649. .4395
```

```
                /gene="wbhC"
CDS             3649. .4395
                /gene="wbhC"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="glycosyl transferase"
                /protein_id="BAA77726.1"
                /db_xref="GI:4867917"
                /translation="MKISVITVTYNNAEGLEKTLSSLSILKIKPFEIIIVDGGSTDGT
                NRVISRFTSMNITHVYEKDEGIYDAMNKGRMLAKGDLIHYLNAGDSVIGDIYKNIKEP
                CLIKVGLFENDKLLGFSSITHSNTGYCHQGVIFPKNHSEYDLRYKICADYKLIQEVFP
                EGLRSLSLITSGYVKYDMGGVSSKKRILRDKELAKIMFEKNKKNLIKFIPISIIKILF
                PERLRRVLRKMQYICLTLFFMKNSSPYDNE"
gene            4388. .5773
                /gene="wzx"
CDS             4388. .5773
                /gene="wzx"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="flippase"
                /protein_id="BAA77727.1"
                /db_xref="GI:4867918"
                /translation="MNKIKKILKFCTLKKYDTSSALGREQERYRIISLSVISSLISKI
                LSLLSLILTVSLTLPYLGQERFGVWMTITSLGAALTFLDLGIGNALTNRIAHSFACGK
                NLKMSRQISGGLTLLAGLSFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKVFV
                IIFGLGIYSNGVQKVYMGIQKAYISNIVNAIFILLSIITLVISSKLHAGLPVLIVSTL
                GIQYISGIYLTINLIIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSGDN
                FIISITLGVTYVAVFSITQRLFQISTVPLTIYNIPLWAAYADAHARNDTQFIKKTLRT
                SLKIVGISSFLLAFILVVFGSEVVNIWTEGKIQVPRTFIIAYALWSVIDAFSNTFASF
                LNGLNIVKQQMLAVVTLILIAIPAKYIIVSHFGLTVMLYCFIFIYIVNYFIWYKCSFK
                KHIDRQLNIRG"
gene            5776. .6870
                /gene="wbhD"
CDS             5776. .6870
                /gene="wbhD"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="perosamine synthetase"
                /protein_id="BAA77728.1"
                /db_xref="GI:4867919"
                /translation="MKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKFAE
                QNHVQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFVDS
                DNETWQMSVSDIEQKITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEAFG
                SKYKGKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHRQY
                WHDVIGYNYRMTNICAAIGLAQLEQADDFISRKREIADIYKKNINSLVQVHKESKDVF
                HTYWMVSILTRTAEEREELRNHLADKLIETRPVFYPVHTMPMYSEKYQKHPIAEDLGW
                RGINLPSFPSLSNEQVIYICESINEFYSDK"
gene            6895. .8109
                /gene="wbhE"
CDS             6895. .8109
                /gene="wbhE"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="glycosyl transferase"
                /protein_id="BAA77729.1"
                /db_xref="GI:4867920"
                /translation="MKIALNSDGFYEWGGGIDFIKYILSILETKPEICIDILLPRNDI
                HSLIREKAFPFKSILKAILKRERPRWISLNRFNEQYYRDAFTQNNIETNLTFIKSKSS
                AFYSYFDSSDCDVILPCMRVPSGNLNKKAWIGYIYDFQHCYYPSFFSKREIDQRNVFF
                KLMLNCANNIIVNAHSVITDANKYVGNYSAKLHSLPFSPCPQLKWFADYSGNIAKYNI
                DKDYFIICNQFWKHKDHATAFRAFKIYTEYNPDVYLVCTGATQDYRFPGYFNELMVLA
                KKLGIESKIKILGHIPKLEQIELIKNCIAVIQPTLFEGGPGGGVTFDAIALGKKVILS
                DIDVNKEVNCGDVYFFQAKNHYSLNDAMVKADESKIFYEPTTLIELGLKRRNACADFL
                LDVVKQEIESRS"
gene            8129. .9247
                /gene="gmd"
CDS             8129. .9247
                /gene="gmd"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose 4,6-dehydratase"
                /protein_id="BAA77730.1"
                /db_xref="GI:4867921"
                /translation="MTKVALITGVTGQDGSYLAEFLLDKGYEVHGIKRRASSFNTERI
                DHIYQDPHGSNPNFHLHYGDLTDSSNLTRILKEVQPDEVYNLAAMSHVAVSFESPEYT
                ADVDAIGTLRLLEAIRFLGLENKTRFYQASTSELYGLVQEIPQKESTPFYPRSPYAVA
                KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRGLANIAQGLESCLYL
                GNMDSLRDWGHAKDYVRMQWLMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLGIKMSF
                VGKGIEEKGIVDSVEGQDAPGVKPGDVIVAVDPRYFRPAEVDTLLGDPSKANLKLGWR
                PEITLAEMISEMVAKDLEAAKKHSLLKSHGFSVSLALE"
gene            9250. .10215
                /gene="wbhF"
CDS             9250. .10215
                /gene="wbhF"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="GDP-L-fucose pathway enzyme"
                /protein_id="BAA77731.1"
                /db_xref="GI:4867922"
                /translation="MNKQRIFIAGHQGMVGSAITRRLKQRDDVELVLRTRDELNLLDS
                SAVLDFFSSQKIDQVYLAAAKVGGILANSSYPADFIYENIMIEANVIHAAHKNNVNKL
                LFLGSSCIYPKLAHQPIMEDELLQGKLEPTNEPYAIAKIAGIKLCESYNRQFGRDYRS
                VMPTNLYGPNDNFHPSNSHVIPALLRRFHDAVENNSPNVVVWGSGTPKREFLHVDDMA
                SASIYVMEMPYDIWQKNTKVMLSHINIGTGIDCTICELAETIAKVVGYKGHITFDTTK
                PDGAPRKLLDVTLLHQLGWNHKITLHKGLENTYNWFLENQLQYRG"
gene            10218. .10727
                /gene="wbhG"
CDS             10218. .10727
                /gene="wbhG"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="GDP-L-fucose pathway enzyme"
                /protein_id="BAA77732.1"
                /db_xref="GI:4867923"
                /translation="MFLHSQDFATIVRSTPLISIDLIVENEFGEILLGKRINRPAQGY
                WFVPGGRVLKDEKLQTAFERLTEIELGIRLPLSVGKFYGIWQHFYEDNSMGGDFSTHY
                IVIAFLLKLQPNILKLPKSQHNAYCWLSRAKLINDDDVHYNCRAYFNNKTNDAIGLDN
                KDIICLMRQ"
gene            10739. .12157
                /gene="manC"
CDS             10739. .12157
                /gene="manC"
                /note="putative"
                /codon_start=1
                /transl_table=11
```

```
Query Match             100.0%;  Score 1107;  DB 1;  Length 31482;
Best Local Similarity   100.0%;  Pred. No. 3.4e-168;
Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7003 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 7062

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7063 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 7122

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7123 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 7182

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7183 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 7242

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7243 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 7302
```

```
Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7303 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 7362

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7363 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 7422

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7423 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 7482

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7483 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 7542

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7543 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 7602

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7603 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 7662

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7663 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 7722

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7723 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 7782

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7783 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 7842

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7843 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 7902

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7903 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 7962

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7963 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 8022

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8023 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 8082

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
              ||||||||||||||||||||||||||
Db       8083 AAACAAGAAATTGAATCCCGATCTTAA 8109
```

RESULT 8
AE005174_28/c
WPCOMMENT
Sequence split into 56 fragments LOCUS AE005174 Accession AE005174

Fragment Name	Begin	End
AE005174_00	1	110000
AE005174_01	100001	210000
AE005174_02	200001	310000
AE005174_03	300001	410000
AE005174_04	400001	510000
AE005174_05	500001	610000
AE005174_06	600001	710000
AE005174_07	700001	810000
AE005174_08	800001	910000
AE005174_09	900001	1010000
AE005174_10	1000001	1110000
AE005174_11	1100001	1210000
AE005174_12	1200001	1310000
AE005174_13	1300001	1410000
AE005174_14	1400001	1510000
AE005174_15	1500001	1610000
AE005174_16	1600001	1710000
AE005174_17	1700001	1810000
AE005174_18	1800001	1910000
AE005174_19	1900001	2010000
AE005174_20	2000001	2110000
AE005174_21	2100001	2210000
AE005174_22	2200001	2310000
AE005174_23	2300001	2410000
AE005174_24	2400001	2510000
AE005174_25	2500001	2610000
AE005174_26	2600001	2710000
AE005174_27	2700001	2810000
AE005174_28	2800001	2910000
AE005174_29	2900001	3010000
AE005174_30	3000001	3110000
AE005174_31	3100001	3210000
AE005174_32	3200001	3310000
AE005174_33	3300001	3410000
AE005174_34	3400001	3510000
AE005174_35	3500001	3610000
AE005174_36	3600001	3710000
AE005174_37	3700001	3810000
AE005174_38	3800001	3910000
AE005174_39	3900001	4010000
AE005174_40	4000001	4110000
AE005174_41	4100001	4210000
AE005174_42	4200001	4310000
AE005174_43	4300001	4410000
AE005174_44	4400001	4510000
AE005174_45	4500001	4610000
AE005174_46	4600001	4710000
AE005174_47	4700001	4810000
AE005174_48	4800001	4910000
AE005174_49	4900001	5010000
AE005174_50	5000001	5110000
AE005174_51	5100001	5210000
AE005174_52	5200001	5310000
AE005174_53	5300001	5410000
AE005174_54	5400001	5510000
AE005174_55	5500001	5528445

Continuation (29 of 56) of AE005174 from base 2800001 (AE005174 Escherichia coli O157:H7

Query Match 100.0%; Score 1107; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred. No. 2.2e-168;
Matches 1107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52202 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 52143

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52142 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 52083

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52082 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 52023

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52022 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 51963

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51962 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 51903

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
```

```
Db      51902 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 51843

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420

Db      51842 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 51783

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480

Db      51782 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 51723

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540

Db      51722 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 51663

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600

Db      51662 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 51603

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660

Db      51602 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 51543

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720

Db      51542 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 51483

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780

Db      51482 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 51423

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840

Db      51422 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 51363

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900

Db      51362 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 51303

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960

Db      51302 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 51243

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020

Db      51242 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 51183

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080

Db      51182 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 51123

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107

Db      51122 AAACAAGAAATTGAATCCCGATCTTAA 51096


RESULT 9
BA000007_27/c
WPCOMMENT
  Sequence split into 55 fragments  LOCUS BA000007 Accession BA000007
     Fragment Name      Begin     End
     BA000007_00            1    110000
     BA000007_01       100001    210000
     BA000007_02       200001    310000
     BA000007_03       300001    410000
     BA000007_04       400001    510000
     BA000007_05       500001    610000
     BA000007_06       600001    710000
     BA000007_07       700001    810000
     BA000007_08       800001    910000
     BA000007_09       900001   1010000
     BA000007_10      1000001   1110000
     BA000007_11      1100001   1210000
     BA000007_12      1200001   1310000
     BA000007_13      1300001   1410000
     BA000007_14      1400001   1510000
     BA000007_15      1500001   1610000
     BA000007_16      1600001   1710000
     BA000007_17      1700001   1810000
     BA000007_18      1800001   1910000
     BA000007_19      1900001   2010000
     BA000007_20      2000001   2110000
     BA000007_21      2100001   2210000
     BA000007_22      2200001   2310000
     BA000007_23      2300001   2410000
     BA000007_24      2400001   2510000
     BA000007_25      2500001   2610000
     BA000007_26      2600001   2710000
     BA000007_27      2700001   2810000
     BA000007_28      2800001   2910000
     BA000007_29      2900001   3010000
     BA000007_30      3000001   3110000
     BA000007_31      3100001   3210000
     BA000007_32      3200001   3310000
     BA000007_33      3300001   3410000
     BA000007_34      3400001   3510000
     BA000007_35      3500001   3610000
     BA000007_36      3600001   3710000
     BA000007_37      3700001   3810000
     BA000007_38      3800001   3910000
     BA000007_39      3900001   4010000
     BA000007_40      4000001   4110000
     BA000007_41      4100001   4210000
     BA000007_42      4200001   4310000
     BA000007_43      4300001   4410000
     BA000007_44      4400001   4510000
     BA000007_45      4500001   4610000
     BA000007_46      4600001   4710000
     BA000007_47      4700001   4810000
     BA000007_48      4800001   4910000
     BA000007_49      4900001   5010000
     BA000007_50      5000001   5110000
     BA000007_51      5100001   5210000
     BA000007_52      5200001   5310000
     BA000007_53      5300001   5410000
     BA000007_54      5400001   5498450
Continuation (28 of 55) of BA000007 from base 2700001 (BA000007 Escherichia coli O157:H7

  Query Match           100.0%;  Score 1107;  DB 1;  Length 110000;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-168;
  Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60

Db      81988 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 81929

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120

Db      81928 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 81869

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180

Db      81868 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 81809

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240

Db      81808 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 81749

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300

Db      81748 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 81689

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360

Db      81688 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 81629
```

```
Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81628 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 81569

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81568 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 81509

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81508 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 81449

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81448 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 81389

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81388 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 81329

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81328 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 81269

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81268 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 81209

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81208 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 81149

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81148 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 81089

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81088 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 81029

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81028 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 80969

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80968 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 80909

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
              ||||||||||||||||||||||||||||
Db      80908 AAACAAGAAATTGAATCCCGATCTTAA 80882
```

```
RESULT 10
AY730593
LOCUS       AY730593               14323 bp    DNA     linear   BCT 25-OCT-2004
DEFINITION  Salmonella enterica subsp. enterica serovar Urbana UDP-glucose
            pyrophosphorylase (galF) gene, partial cds; O-antigen gene cluster,
            complete sequence; and 6-phosphogluconate dehydrogenase (gnd) gene,
            partial cds.
ACCESSION   AY730593 AY012194
VERSION     AY730593.1  GI:54401420
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Urbana
  ORGANISM  Salmonella enterica subsp. enterica serovar Urbana
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 14323)
  AUTHORS   Jensen,S.O. and Reeves,P.R.
  TITLE     Molecular evolution of the GDP-mannose pathway genes (manB and
            manC) in Salmonella enterica
  JOURNAL   Microbiology (Reading, Engl.) 147 (Pt 3), 599-610 (2001)
   PUBMED   11238967
REFERENCE   2  (bases 1 to 14323)
  AUTHORS   Samuel,G., Hogbin,J.P., Wang,L. and Reeves,P.R.
  TITLE     Relationships of the Escherichia coli O157, O111, and O55 O-Antigen
            Gene Clusters with Those of Salmonella enterica and Citrobacter
            freundii, Which Express Identical O Antigens
  JOURNAL   J. Bacteriol. 186 (19), 6536-6543 (2004)
   PUBMED   15375135
REFERENCE   3  (bases 1 to 14323)
  AUTHORS   Jenkins,S.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-2000) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
REFERENCE   4  (bases 1 to 14323)
  AUTHORS   Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2004) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
  REMARK    Sequence update by submitter
COMMENT     On Oct 25, 2004 this sequence version replaced gi:11992707.
FEATURES             Location/Qualifiers
     source          1. .14323
                     /organism="Salmonella enterica subsp. enterica serovar
                     Urbana"
                     /mol_type="genomic DNA"
                     /strain="M284"
                     /serotype="30"
                     /serovar="Urbana"
                     /sub_species="enterica"
                     /db_xref="taxon:260677"
     gene            <1. .852
                     /gene="galF"
     CDS             <1. .852
                     /gene="galF"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-glucose pyrophosphorylase"
                     /protein_id="AAV34505.1"
                     /db_xref="GI:54401421"
                     /translation="MHMLPATKAIPKEMLPIVDKPMIQYIVDEIVAAGIKEIVLVTHA
                     SKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNVRQAQPLGLGHSILC
                     ARPVVGDNPFIVVLPDIIIDDATADPLRYNLAAMVARFNETGRSQVLAKRMKGDLSEY
                     SVIQTKEPLDNEGKVSRIVEFIEKPDQPQTLDSDLMAVGRYVLSADIWAELERTEPGA
                     WGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKYGLRNLKEGAKFRKS
                     IEQLLHE"
     gene            1220. .2002
                     /gene="wbdN"
     CDS             1220. .2002
                     /gene="wbdN"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAV34506.1"
                     /db_xref="GI:54401422"
                     /translation="MNREIVSVIMPVYNGSHTIVDSIESVLNQTYKDIKLYVIDDCST
                     DDTAKLLTVNYANNKNVKIIISSKNKGVAASRNLGIKESKGKYIAFCDADDLWENNKI
                     EKQINLLEHNNINVVCSNYYIMDFNKKIIGETHSPVTISYKKMLMKNYVGNLTGVYNA
                     EKLGKFYQKKVGHEDYLMWLEIIKKSNFAICLQENLAFYRVSNNSLSASKLVAAKWTW
                     NIYRECLGLSLPKSLYYFSLYAANGVIKKIGLINSDNSKAKE"
     gene            2255. .3157
                     /gene="wzy"
     CDS             2255. .3157
                     /gene="wzy"
                     /codon_start=1
                     /transl_table=11
                     /product="O-antigen polymerase"
                     /protein_id="AAV34507.1"
                     /db_xref="GI:54401423"
                     /translation="MEIMAIILFAKEGIPIFDADPGGAKLKIADGNGIYIRYIKYFGN
                     IVTFTLLLLLNKQKVKQKIIAIIYFLTIALFGYRSELVLLFLQYYIIMSIVNEGGNKY
                     KFLNKIFIFIFMGGVIGSLFYFSLGQDGAVNNSTENIDRIINRLTIEQVESVPYIISE
                     SLEYDLFPTPELGKEINAIFNRILGEKDQNLFYGERLHKKVFGDMGANFLSVTTYGAE
                     LLAFFGPLCIFLIPIGLCIPFAMLKKISNTHSTVNGAIYSFLIMILLQYLIAGNASAF
                     FFGPFITAIVMSIPLMILHDILKR"
```

```
gene            3163. .3894
                /gene="wbdO"
CDS             3163. .3894
                /gene="wbdO"
                /codon_start=1
                /transl_table=11
                /product="glycosyltransferase"
                /protein_id="AAV34508.1"
                /db_xref="GI:54401424"
                /translation="MNTSVITVTYNNAEGLNKTLESLCLLKVKPYEIIIIDANSIDNT
                KAIVKKYSDMNIIYICEKDNGIYDAMNKGLRLARGKLLHYLNAGDCVSGEIYADLNGP
                CLLKVGLFDKDNLIGFSSLTHSNTGYCHQGIIFPSKHPIYDLKYKLCADYKLIQMVFP
                MGLKQLPLISNGFIKYDMGGVSANKRFLRDKELLQIMLETNKKTLIKFIPLMVIKMVF
                PKKIRRILRRVNYYFSLILTNKKNV"
gene            3896. .5287
                /gene="wzx"
CDS             3896. .5287
                /gene="wzx"
                /codon_start=1
                /transl_table=11
                /product="O-antigen flippase"
                /protein_id="AAV34509.1"
                /db_xref="GI:54401425"
                /translation="MIKEKIKILFKYCTLRKYNTTTDYGRQQERYRIISLSIISNFIS
                KLLALLSLILTVSLTLPYLGQDRFGIWMTIASFSAALTFLDLGIGNALTNRIAHSFAR
                GKKIRMIRQISGGLTLLAILAFIIIAIFLTIANFVNWGSLLKGINDNTHDEVQFAIKL
                FAIIFGIGIFSNGVQRIYMGMQKAYISYISNAFFILISIISLFISSRLHAGIPMLILS
                TIGVQYLSGLSLTILLIYKKLITLNKIRINVKKETPYLISNGFFFFILQMGTLATWSG
                DNFIISTTLGVTYVSILSITQRLFQISTVPLTIYNAPLWAAYADAHARHDVDFIRKTL
                LGSLKIVGISSTTLALILVIFGQSIIAIWTGEELRVPQSLIVAYAVWSIIDAYSNTLS
                SFLNGLNIIKQQVFAVVTLILISIPSKYFIVSHFGLTAMLSCFIILYVTNYLIWYKCS
                FKKSIHLQLNIRG"
gene            5290. .6384
                /gene="per"
CDS             5290. .6384
                /gene="per"
                /codon_start=1
                /transl_table=11
                /product="perosamine synthetase"
                /protein_id="AAV34510.1"
                /db_xref="GI:54401426"
                /translation="MEYIPVYQPALSGKEKEYVNDCIDSTWISSKGSYIQRFEKKFAE
                QNNIQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPVFVDS
                DLQTWQISVDDIKNKITSRTKAIMCVHLYGHPCDMSPILEIAKVNNLFIIEDCAEAFG
                SKYKDKFVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAAHRQY
                WHDVIGYNYRMTNICAAIGLAQLEQAQLFIARKREIAEIYKNNLTPLIEVHQEGNDIF
                HTYWMVSILARNAEEREVLRSFLSEKLTETRPLFYPVHTMPMYSDKYQKHPVAENLGW
                RGINLPSFPTLSNEQVIYICNAINEFYKDK"
gene            6409. .7623
                /gene="wbdP"
CDS             6409. .7623
                /gene="wbdP"
                /codon_start=1
                /transl_table=11
                /product="glycosyltransferase"
                /protein_id="AAV34511.1"
                /db_xref="GI:54401427"
                /translation="MKITLNTDGFNEWGGGVDFIKYLLRTLEIEQDIFLDVLLPREDI
                SSLLKRNIFPVKNMIKAILSGQTPRWKKLEGFNEQYYKHAFAEITEQSKLSFITGKSS
                AYFSHYIKSDYDIILPCMRTPAVNMKNRAWLGYIYDFQHCYYPSFFSEREIAKRNLFF
                KRMLDNANNVIVNSHAVISDAKKFVGDFSANLHALPFSPCPELNWLNDNSDLIDKYAI
                DKDYFIICNQFWKHKDHSTAFKAFKKYIEINPDVYLVCTGATQDYRFPGYFDELMCLL
                KKLGISEKVKILGHIPKLEQIELIKNSIAVIQPTLFEGGPGGGITFDAVALGKKIVLS
                DIEVNKDVDCGHVYFFHSQNEYSLTEALVKATAVDHYYDPHVLLESGIKRRKKCADFI
                LNVISKEIESRI"
gene            7647. .8765
                /gene="gmd"
CDS             7647. .8765
                /gene="gmd"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose dehydratase"
                /protein_id="AAV34512.1"
                /db_xref="GI:54401428"
                /translation="MSKVALITGVTGQDGSYLAEFLLEKGYEVHGIKRRASSFNTERV
                DHIYQDPHSCNPKFHLHYGDLTDASNLTRILQEVQPDEVYNLGAMSHVAVSFESPEYT
                ADVDAMGTLRLLEAIRFLGLEKKTRFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
                KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRAIANIAQGLESCLYL
                GNMDSLRDWGHAKDYVRMQWMMLQQEQPEDFVIATGVQYSVRQFVELAAAQLGIKLRF
                EGEGINEKGIVVSVTGHDAPGVKPGDVIVAVDPRYFRPAEVETLLGDPSKAHEKLGWK
                PEITLSEMVSEMVANDLEAAKKHSLLKSHGFNVNLSLE"
gene            8768. .9733
                /gene="fcl"
CDS             8768. .9733
                /gene="fcl"
                /codon_start=1
                /transl_table=11
                /product="fucose synthetase"
                /protein_id="AAV34513.1"
                /db_xref="GI:54401429"
                /translation="MNKQRIFVAGHRGMVGSAIVRQLAQRGDVELVLRTRDELDLLDG
                RAVQAFFAGTGIDQVYLAAAKVGGIVANNTYPADFIYENMMIESNIIHAAHLHNVNKL
                LFLGSSCIYPKLARQPMAESELLQGTLEPTNEPYAIAKIAGIKLCESYNRQYGRDYRS
                VMPTNLYGPHDNFHPDNSHVIPALLRRFHEAAQSHAPEVVVWGSGTPMREFLHVDDMA
                AASIHVMELSREVWQENTAPMLSHINVGTGVDCTISELAQTIAKVVGYQGRVVFDAAK
                PDGTPRKLLDVTRLHQLGWYHEISLEAGLAGTYQWFLENQQRFRG"
gene            9736. .10236
                /gene="gmm"
CDS             9736. .10236
                /gene="gmm"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose mannosylhydrolase"
                /protein_id="AAV34514.1"
                /db_xref="GI:54401430"
                /translation="MFLRQEDFAAVVRATPLISLDFIVENGQGEILLGQRLNRPAQGY
                WFVPGGRVRKDETLEAAFARLTQAELGVRLPLAAGTFYGVWQHFYDDNFSGEDFSTHY
                IVLGFRLRVAESDLLLPDAQHGSYRWLTPEQLLASDNVHDNSRAYFQNAPYSVIGLDK
                KDVKYV"
```

```
Query Match          51.8%;  Score 573.2;  DB 1;  Length 14323;
Best Local Similarity  69.9%;  Pred. No. 1.2e-82;
Matches  773;  Conservative    0;  Mismatches  333;  Indels    0;  Gaps    0;

Qy          2 TTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAAA 61
              ||||||| |||||  | ||||||  |||  |  |||      ||    || ||  | || |
Db       6518 TTCTTTTGCCGAGGGAAGATATAAGTTCCTTGCTAAAGAGGAATATCTTCCCAGTAAAGA 6577

Qy         62 GTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTTA 121
              |||  | || ||  |||||||    |  |||  ||||||||||        ||  |  ||||
Db       6578 ATATGATCAAGGCTATTTTAAGCGGCCAAACTCCTCGATGGAAAAAGCTAGAAGGGTTTA 6637

Qy        122 ATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTTA 181
              |||| || ||||| |   |||| ||| |  ||| || |    || | ||  | |  || |
Db       6638 ATGAACAGTACTACAAGCATGCTTTTGCTGAAATTACTGAGCAGTCTAAGTTAAGTTTCA 6697

Qy        182 TTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATTC 241
              |||   | || ||  ||||| |  | |||| |||    |   ||| | |||| ||||||
Db       6698 TTACTGGCAAAAGTTCTGCTTACTTTTCACATTACATAAAATCCGACTATGATATTATTC 6757

Qy        242 TTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTT 301
              |||| || ||||||  |||| | | |||| |||| ||    |||||| | || |||||||
Db       6758 TTCCCTGTATGCGTACTCCTGCAGTAAATATGAAAAACCGCGCATGGCTAGGATATATTT 6817

Qy        302 ATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGA 361
              |||| |||||||| ||||| || ||||| ||||||||  |  | || |||| | |||| |
Db       6818 ATGATTTTCAACATTGTTATTACCCTTCTTTTTTTAGCGAAAGGGAGATAGCTAAAAGAA 6877

Qy        362 ATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAG 421
              ||  |||||||||||   |||||  ||   ||||| ||| | ||||| ||| | ||| | |
Db       6878 ATTTGTTTTTTAAAAGAATGCTTGATAATGCTAATAATGTCATTGTCAATTCCCATGCTG 6937

Qy        422 TTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCAT 481
              | ||| | |||||||| || | ||||||  ||| |||||||||  |||||  |  | || |
Db       6938 TCATTTCAGATGCAAAAAAGTTTGTTGGTGATTTTTCTGCAAATTTACATGCATTACCTT 6997
```

```
Qy       482 TTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATA 541
Db      6998 TTAGCCCGTGTCCTGAGCTAAATTGGCTTAATGATAACTCAGATCTTATTGATAAATATG 7057

Qy       542 ATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAA 601
Db      7058 CTATTGATAAAGATTACTTTATAATATGTAATCAGTTCTGGAAGCATAAAGATCATTCAA 7117

Qy       602 CTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGCA 661
Db      7118 CGGCTTTTAAAGCTTTTAAGAAATATATAGAAATAAACCCTGATGTATATTTAGTATGCA 7177

Qy       662 CGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCAA 721
Db      7178 CAGGTGCGACGCAAGATTATAGATTTCCTGGATATTTTGACGAACTTATGTGTTTGCTTA 7237

Qy       722 AAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAAA 781
Db      7238 AAAAATTGGGTATATCTGAGAAAGTAAAAATATTAGGGCATATTCCAAAATTAGAGCAAA 7297

Qy       782 TTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCTG 841
Db      7298 TAGAACTAATAAAGAATAGTATTGCTGTGATACAACCCACTTTGTTTGAGGGGGGGCCGG 7357

Qy       842 GAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATAG 901
Db      7358 GAGGAGGAATTACATTTGATGCAGTTGCATTAGGGAAAAAGATAGTATTATCTGATATAG 7417

Qy       902 ATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTATT 961
Db      7418 AGGTTAATAAAGATGTTGATTGTGGGCATGTTTATTTTTTCCATTCTCAAAATGAATACT 7477

Qy       962 CATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACTC 1021
Db      7478 CTTTAACTGAAGCTTTAGTTAAAGCCACTGCAGTTGATCATTATTATGACCCACATGTCT 7537

Qy      1022 TGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTGA 1081
Db      7538 TACTTGAATCTGGGATAAAAAGGCGCAAAAAATGTGCTGACTTTATTTTAAATGTAATAA 7597

Qy      1082 AACAAGAAATTGAATCCCGATCTTAA 1107
Db      7598 GTAAAGAAATAGAATCACGTATTTGA 7623
```

RESULT 11
AY730592
LOCUS AY730592 14224 bp DNA linear BCT 02-NOV-2004
DEFINITION Citrobacter freundii serotype 90 UDP-glucose pyrophosphorylase (galF) gene, partial cds; O-antigen gene cluster, complete sequence; and 6-phosphogluconate dehydrogenase (gnd) gene, partial cds.
ACCESSION AY730592
VERSION AY730592.2 GI:55228555
KEYWORDS .
SOURCE Citrobacter freundii
ORGANISM Citrobacter freundii
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Citrobacter.
REFERENCE 1 (bases 1 to 14224)
AUTHORS Samuel,G., Hogbin,J.P., Wang,L. and Reeves,P.R.
TITLE Relationships of the Escherichia coli O157, O111, and O55 O-Antigen Gene Clusters with Those of Salmonella enterica and Citrobacter freundii, Which Express Identical O Antigens
JOURNAL J. Bacteriol. 186 (19), 6536-6543 (2004)
PUBMED 15375135
REFERENCE 2 (bases 1 to 14224)
AUTHORS Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2004) School of Molecular and Microbial Biosciences, University of Sydney, Sydney, NSW 2006, Australia
REFERENCE 3 (bases 1 to 14224)
AUTHORS Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-2004) School of Molecular and Microbial Biosciences, University of Sydney, Sydney, NSW 2006, Australia
REMARK Sequence update by submitter
COMMENT On Nov 2, 2004 this sequence version replaced gi:54401406.
FEATURES Location/Qualifiers

```
     source          1. .14224
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /strain="M1972"
                     /serotype="90"
                     /db_xref="taxon:546"
     gene            <1. .773
                     /gene="galF"
     CDS             <1. .773
                     /gene="galF"
                     /codon_start=3
                     /transl_table=11
                     /product="UDP-glucose pyrophosphorylase"
                     /protein_id="AAV34492.2"
                     /db_xref="GI:55228556"
                     /translation="FVIVAAGIKEIVLVTHASKNAVENHFDTSYELESLLEQRVKRQL
                     LAEVQSICPPGVTIMNVRQAQLLGLGHSILCARPIIGDNPFVVVLPDVVLDTASADPL
                     RYNLAAMVARFNETGRSQVLAKRMEGDLSEYSVIQTKDPLDNEGKVSRIVEFIEKPDQ
                     PQTLDSDLMAVGRYVLSADIWAELEKTAPGAWGRIQLTDAISEPAKKQSVDAILMTGE
                     SYDCGKKMGYMQAFVKYGLRNLKEGPKFRKSIEKLLAM"
     gene            1104. .1853
                     /gene="wbdN"
     CDS             1104. .1853
                     /gene="wbdN"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAV34493.1"
                     /db_xref="GI:54401408"
                     /translation="METESVSIIMPVYNGASTIVQSIESILNQTYKDIRLYIIDDCST
                     DETSILLQNLYISNDRVELIRNIKNKGVAESRNIGIEQSRGKYIAFCDADDLWGETKL
                     EKQLDILESEKVDVVCTNYYLIDLEDNIIGSVCSPGIITYKTMLGKNHIGNLTGIYNV
                     KKIGKVYQKKIGHEDYLMWLEILRKTEQAICIQENLAKYRVSKKSLSANKLIAAQWTW
                     NIYRKQLKIPIIPALYFFSQYAVQGLLKKII"
     gene            1890. .3047
                     /gene="wzy"
     CDS             1890. .3047
                     /gene="wzy"
                     /codon_start=1
                     /transl_table=11
                     /product="O-antigen polymerase"
                     /protein_id="AAV34494.1"
                     /db_xref="GI:54401409"
                     /translation="MKKSVGIISFFLFSFYVLQLYGLLIDVRIKDVNIQIIAGGLVLF
                     QILCVLFYALFIRKNLSYQSGLFFNSKKIIIIFLLTFVTIELAAVILFLKEGIPIFDA
                     DPGGAKLRIADGNGIYIRYIKYFGNIVCFSLCLLLKDRKILHYSVVIIYFSTISLFGY
                     RSELIMLMIQYYILMSIINNNKVKHSMLRKVVILFSISCIVIGLFYFSLGQEGQENNY
                     ASNVNRIFNRLTIEQVESVPYIINESMKYDLFPTSEIGKEINAIANRFQGIKDNSLFY
                     GERLHKDIFGDMGENFLSVTTFGAELLAFFGPFCFFLIPLGIYIPYFLLIKMSQTKNT
                     VLCSAYSYSVVIFLQYLIAGNFSAFFFGPILTIVIMCIPLVLLHDVLKRLN"
     gene            3047. .3781
                     /gene="wbdO"
     CDS             3047. .3781
                     /gene="wbdO"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAV34495.1"
                     /db_xref="GI:54401410"
                     /translation="MKISVVTVTYNNGNGLRKTLQSLSSIKTKPCEIIIIDAESKDNT
                     KSVVDEYFSKMNIIYINEKDNGIYDAMNKGHKIANGDYIHYLNAGDCVSGDVYANVTG
                     TCLIKVGLFNERDSFLGFSNLTHSNTGYCHQGVIFPKEHLDYDLKYKLCADYDLLQKT
                     FPEGLASLPLITDGYIKYDMSGVSANKTLLRDKELLAIMWKNNKRIIPKFMIYVLFKT
                     MIPKGCRRFLRKLKYSISVFSLNGNK"
     gene            3789. .5180
                     /gene="wzx"
     CDS             3789. .5180
```

/gene="wzx"
/codon_start=1
/transl_table=11
/product="O-antigen flippase"
/protein_id="AAV34496.1"
/db_xref="GI:54401411"
/translation="MNNLLRKIYSVLKYCTLKKFDTSTDYGRQQERYRIISLSIIANL
ISKFLSLLTLVLTVSLTLPYLGPARFGVWMTIASLSAALTFLDLGIGNALTNRIAHAC
ANGERTEVIKKTSGGIALLLLVSICITCVLYIICSVIDWRIFIKGINDNILGEVQRAI
KLFVVIFGISILSNGIQKIYMGMQKAYITYFFNSLLSLISIFCLIYYSSLHATIPVLI
LISSGIPLFTGMLLMILLIKDGFVSLNEIFHNMRNEAPHLISTGFFFFILQIGTLATW
SGDNFIISSTLGVAFVAILSLTQRLFQISTVPLTIYNTPLWVAYADAHARKDTHFIKK
TLLASFKIVGFSSLIMATFLILFGRDIISVWTGDKIDIPQSFIIAYAIWSIIDAVSNT
LASFLNGLGIVKQQMFAVVSLVIISIPAKYFIVSHFGLTAMLSCFIAIYIVNYIVWYT
ISFKKCIYSQLSI"

gene 5188. .6285
/gene="per"

CDS 5188. .6285
/gene="per"
/codon_start=1
/transl_table=11
/product="perosamine synthetase"
/protein_id="AAV34497.1"
/db_xref="GI:54401412"
/translation="MKYIPVYEPSLTGKEKEYVNDCLDSTWISSKGSYIQRFETEFSK
HISAQYATTVSNGTVAIHLALIALGISEGDEVIVPTLTYIASVNAINYTGAIPVFVDS
DAETWQVSIEDIRNKITNKTKAIMCVHLYGQPCDMDAILEIANEFNLFVIEDCAEAFG
SKYKDKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYERCVHYKGQGLAAHRQY
WHDVIGYNYRMTNICAAIGLAQLEQAEGFINRKRNIAEIYKKELFNSPVKVHQESDNT
FHTYWMVSILTRTAEEREELRNYLSENFIETRPLFYPVHTMPMYSEKYQKHPVAEDLG
WRGMNLPSYPGMTDEQVIYICNKIKCFYNEK"

gene 6310. .7524
/gene="wbdP"

CDS 6310. .7524
/gene="wbdP"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAV34498.2"
/db_xref="GI:55228557"
/translation="MKITINTDGFNEWGGGVDFIKYLLLILSTKDDVFTDIIVPKDDF
ASLFRRNVFPIKNAIKSLGKGKFPRWVKKDGFDEGYYHNAFSGLVDKSKLTFINSNSN
SYFDYHAKSDNDIIFPCMRVPTNSVCDSPWIGYIYDFQHCYYPSFFSAREIKKRELYF
REMLSSANNVIVNANSVIADAKNFIGDFSADLHALPFSPCPQQIWLDDNSNLVDKYAI
SKDYFIVCNQFWKHKNHVVVFKAFKEYIKVNPEAYLVCTGGIQDYRFPEYFGELKALI
QKLGISSKIKILGHIPKLEQVELIKKSIAVIQPTLFEGGPGGGATYDAVALGKKVILS
DIDVNREVNCGDVIFFNPQDEYSLAKALFKAEEEKSVKENGDLMKLGLKRREKCADFI
MSVIRKEIESRN"

gene 7546. .8664
/gene="gmd"

CDS 7546. .8664
/gene="gmd"
/codon_start=1
/transl_table=11
/product="GDP-mannose dehydratase"
/protein_id="AAV34499.1"
/db_xref="GI:54401414"
/translation="MSKVALITGVTGQDGSYLAEFLLEKGYEVHGIKRRASSFNTGRV
DHIYQDPHLGNPKFNLHYGDLTDTSNLTRILKEVQPDEVYNLGAMSHVAVSFESPEYT
ADVDAMGTLRLLEAIRFLGLEKKTRFYQASTSELYGLVREIPQKETTPFYPRSPYAVA
KLYAYWITVNYRESYGMYACNGILFNHESPRRGETFVTRKITRAMANIAQGLESCLYL
GNMDSLRDWGHAKDYVRMQWMMLQQEQPEDFVIATGVQYSVRDFVEMSAAQLGIKLSF
EGDGVNEKGIVASVTGHDAPGVKVGDVIVEVDPRYFRPAEVETLLGDPSKAHEKLGWK
PEITLQEMVAEMVASDLEAAKKYSLLKESGFDVNLSLE"

gene 8667. .9632
/gene="fcl"

CDS 8667. .9632
/gene="fcl"
/codon_start=1
/transl_table=11
/product="fucose synthetase"
/protein_id="AAV34500.1"
/db_xref="GI:54401415"
/translation="MKKQRIFVAGHRGMVGTAIVRQLSLRDNVELVLRTRDELNLLDS
SAVQAFFATERIDQVYLAAAKVGGIVANNTYPADFIYENMMIESNIIHAAHLHNVNKL
LFLGSSCIYPKQATQPIAESELLQGTLEPTNEPYAIAKIAGIKLCESYNRQYGRDYRS
VMPTNLYGPHDNFHPSNSHVIPALLRRFHEARENQNAPDVVVWGSGTPMREFLHVDDMA
AASIHVMELDSEIWQEYTQPMLSHINVGTGVDCTIRELAQTIAQVVGYKGKVVFDSSK
PDGTPRKLLDVTRLHSLGWRHSVSLEYGLESTYQWFLENQHWCRG"

gene 9635. .10135
/gene="gmm"

CDS 9635. .10135
/gene="gmm"
/codon_start=1
/transl_table=11
/product="GDP-mannose mannosylhydrolase"
/protein_id="AAV34501.1"
/db_xref="GI:54401416"
/translation="MFLHHDEFATVVRSTPLVSIDLIVENADGEFLLGKRLNRPAQGY
WFVPGGRVQKDERIENAFERLTLAELGLRLPLSVAEFYGVWQHFYDDNFSGPDFSTHY
IVLAFHIRGQESCLCLPDEQHKIWRWFKSEELEEEKDVHVNSRAYFHIENSNVIGLDN
KGLNHV"

gene 10128. .11576
/gene="manC"

CDS 10128. .11576
/gene="manC"
/codon_start=1
/transl_table=11
/product="mannose-1-P guanosyltransferase"
/protein_id="AAV34502.1"
/db_xref="GI:54401417"

Query Match 43.1%; Score 476.6; DB 1; Length 14224;
Best Local Similarity 64.4%; Pred. No. 3.6e-67;
Matches 713; Conservative 0; Mismatches 394; Indels 0; Gaps 0;

```
Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||  |  |||| | | |||| |   |||  |  | |||  ||| |  |||||  | |||
Db       6418 ATAATCGTACCAAAAGATGACTTTGCTTCGTTATTCAGAAGAAACGTTTTTCCCATAAAG 6477

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              | |    | ||  || |   |||  |  ||   ||| |||||  ||  |  | || | |||
Db       6478 AATGCCATCAAGTCATTAGGAAAAGGAAAATTTCCTAGATGGGTTAAAAAAGATGGGTTT 6537

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||    || ||     |||||||  |   |    |    |  | || || || |||
Db       6538 GATGAAGGGTATTACCATAATGCCTTCAGTGGATTGGTTGATAAATCAAAGCTGACTTTT 6597

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              |||||  |||||       | | | || |   |||  || ||      |||  |||| ||||
Db       6598 ATTAATAGTAATTCAAATTCATATTTTGACTATCATGCTAAATCTGATAATGATATTATC 6657

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||  ||||||  | |||||  |         |   | | |   ||||||||||||||||
Db       6658 TTTCCATGCATGAGGGTTCCAACTAATTCAGTTTGTGATAGTCCATGGATTGGTTATATA 6717

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              |||||  |||||  ||  ||  ||  |||||||  ||  ||||||   ||||||||  |  || |
Db       6718 TATGATTTTCAGCATTGCTATTATCCTTCTTTCTTTAGTGCTCGAGAAATTAAAAAACGT 6777

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              |  | | |||||      |||||     | ||  ||  |||  |||| ||||||||  ||||
Db       6778 GAATTATATTTTAGGGAAATGCTTTCCAGTGCGAATAATGTTATAGTTAATGCTAATTCT 6837

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              || ||| | ||||| || ||  |   |||||  ||| |||||||  |  |||||  ||  | |||
Db       6838 GTAATTGCTGATGCTAAAAATTTCATTGGGGATTTTTCTGCAGATTTACATGCTTTACCA 6897

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              || ||||| ||||| |||  ||  |||| | | |||| | ||   |   | |  ||||||
Db       6898 TTCAGTCCTTGCCCACAACAAATATGGCTTGATGATAATTCAAATCTCGTCGATAAATAT 6957

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
                 |||   || ||||| |||||| ||||||||||||||||||||| ||||||  |||| |
Db       6958 GCAATTAGTAAAGATTACTTTATAGTTTGCAATCAATTTTGGAAGCATAAAAATCACGTG 7017
```

```
Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              |  || |  ||||||||    |||| | ||   || || || |  ||||| || ||
Db       7018 GTAGTGTTCAAAGCATTTAAGGAATATATTAAAGTAAACCCCGAAGCATATTTGGTTTGT 7077

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              || |||| || ||| |||||||||||||| | ||| ||   |||  | |  | |||
Db       7078 ACTGGAGGTATTCAGGATTATCGATTCCCGGAATACTTCGGTGAGCTTAAAGCTTTAATT 7137

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||  | ||||||  ||| ||||||| || || || |||||||| |||||||| |||
Db       7138 CAAAAATTGGGAATTTCATCTAAAATTAAAATTTTGGGACATATACCAAAACTTGAGCAA 7197

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              | ||  | ||||||||  |||| || || || ||||||||  | |||||||| || ||
Db       7198 GTGGAGCTGATCAAAAAAAGCATAGCAGTTATTCAACCAACTCTTTTTGAAGGTGGTCCA 7257

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              || |||||||  || | ||| ||  | ||  | |||||||| || ||  |||| || ||
Db       7258 GGTGGGGGGGCTACTTATGATGCGGTAGCTCTCGGGAAAAAGGTGATTTTATCAGATATT 7317

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||| || | |||||||||||| ||||||||   ||||||| |  || || |  | |||
Db       7318 GATGTTAACAGAGAAGTTAATTGTGGTGATGTTATTTTCTTTAATCCACAAGATGAGTAT 7377

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              || ||  | | ||  |  | ||||| || ||    |||  | || | |||  |     |
Db       7378 TCTTTGGCTAAAGCATTATTTAAAGCCGAGGAGGAGAAATCTGTTAAAGAAAATGGTGAT 7437

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              | ||  | || || || ||| ||||  |    ||||| || ||  || |    |||| |
Db       7438 TTAATGAAGTTAGGGCTTAAACGACGTGAAAAATGTGCTGACTTCATTATGAGTGTTATC 7497

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
              | | ||||||| ||||| |||  ||||
Db       7498 AGAAAAGAAATAGAATCGCGAAATTAA 7524


RESULT 12
AC007926/c
LOCUS       AC007926               155204 bp    DNA     linear   HTG 17-JUL-2001
DEFINITION  Trypanosoma brucei chromosome II clone RPCI93-3H15, *** SEQUENCING
            IN PROGRESS ***.
ACCESSION   AC007926
VERSION     AC007926.8  GI:14787210
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 155204)
  AUTHORS   El-Sayed,N.M., Ghedin,E., Song,J., Larkin,C., Wanless,D., Jones,K.,
            Peterson,J., Hou,L., Zhao,H., Mason,T., Militscher,J., Pai,G., Van
            Aken,S., Utterback,T., Khalak,H.G., Gerard,C., Leech,V., Ullu,E.,
            Melville,S., White,O., Adams,M.D., Donelson,J.E. and Fraser,C.M.
  TITLE     Trypanosoma brucei GUTat10.1 RPCI93-3H15 BAC genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155204)
  AUTHORS   El-Sayed,N.M., Khalak,H. and Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Jul 17, 2001 this sequence version replaced gi:12746529.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   155204: contig of 155204 bp in length.
FEATURES            Location/Qualifiers
     source         1. .155204
                    /organism="Trypanosoma brucei"
                    /mol_type="genomic DNA"
                    /isolate="GUTat10.1"
                    /db_xref="taxon:5691"
                    /chromosome="II"
                    /clone="RPCI93-3H15"
ORIGIN

  Query Match             6.9%;  Score 76.6;  DB 14;  Length 155204;
  Best Local Similarity   44.4%;  Pred. No. 0.0017;
  Matches  352;  Conservative    0;  Mismatches  439;  Indels    2;  Gaps    1;

Qy        120 TAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTT 179
              |||| |  |  |  |||  |||     ||   | |||||||   | |  |||  ||
Db      77877 TAATAATAATAATAATAATGATAATGATAATAATAATAATAATAATAATAATAATAATAA 77818

Qy        180 TATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTAT 239
              || ||| | |||       ||     | ||     ||  | ||| ||   ||  | ||  ||
Db      77817 TAATAATAATAATGATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAA 77758

Qy        240 TCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATAT 299
              |  |  |   |   |  |  |   |  |||   ||||| ||        | |  | |||
Db      77757 TAATAATAATAATAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAA 77698

Qy        300 TTATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAG 359
              | || |   | |  |  |  | ||  |    |  || |||     | |   || | |
Db      77697 TAATAATAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 77638

Qy        360 GAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTC 419
              |||     |  |||     |     |||     ||| |||| || |  ||||    ||
Db      77637 TAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 77578

Qy        420 AGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCC 479
              || ||   ||   |||||  ||  |   ||| ||  |   ||     ||   |  |
Db      77577 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 77518

Qy        480 ATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATA 539
              || |         | |   ||              | || |   || |||||  |   || |
Db      77517 TAATAATAATAATAATAATAATAATAATAATAATAATAATGATAATGATAATAATAATAA 77458

Qy        540 TAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGC 599
              ||||| | | ||  || ||  ||  | |   ||| |    |   ||  ||||  || ||
Db      77457 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 77398

Qy        600 AACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATG 659
              | |  |  ||      ||| | | ||| | |  |||||  | ||  | ||
Db      77397 TAATAATAATAATAATAATAATAATAATAATAATAATAATGATAATAATAATAATAATAA 77338

Qy        660 CACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGC 719
              |      | || | |  || ||         |    ||  ||||  | ||||  |  |
Db      77337 TAATAATGATAATGATAATAATAATAATAATAATAATAATAAT--AATGATAATAATAAT 77280

Qy        720 AAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACA 779
              || ||   |  |||   |  || ||| |  ||  ||  |   |||   | | |   |
Db      77279 AATAATAATAATAATGATAATGATAATGATAATGATAATGATAATAATGATAATGATAAT 77220

Qy        780 AATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCC 839
              |||   | |||| |  |      ||    ||||   |  ||   || |   |||  |
Db      77219 AATAATAATAATAATGATAATGATAATGATAATGATAATAATGATAATGATAGGATAGTG 77160

Qy        840 TGGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACAT 899
              | | | | |||      |  |||   | ||    || | | |||  |  ||| ||
Db      77159 TTGTGAGTGTGTATATACGAATAATATAAAAAGAGCAGTAATAATAATAACAATTGAAAT 77100

Qy        900 AGATGTCAATAAA 912
              ||   || |||
Db      77099 TTATCAGAAGAAA 77087


RESULT 13
BX537105
LOCUS       BX537105               225581 bp    DNA     linear   HTG 06-AUG-2003
```

```
DEFINITION  Danio rerio clone DKEY-3L24, 18 unordered pieces.
ACCESSION   BX537105
VERSION     BX537105.2  GI:31559379
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 225581)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 9, 2003 this sequence version replaced gi:31441851.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK3L24
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 217427 bases at least Q40
            Consensus quality: 219130 bases at least Q30
            Consensus quality: 220818 bases at least Q20
            Insert size: 223881; sum-of-contigs
            Insert size: 219000; 7.3% error; agarose-fp
            Quality coverage: 6.82x in Q20 bases; sum-of-contigs Quality
            coverage: 7.43x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    20602: contig of 20602 bp in length
            *    20603    20702: gap of 100 bp
            *    20703    57831: contig of 37129 bp in length
            *    57832    57931: gap of 100 bp
            *    57932    65164: contig of 7233 bp in length
            *    65165    65264: gap of 100 bp
            *    65265    71392: contig of 6128 bp in length
            *    71393    71492: gap of 100 bp
            *    71493    93677: contig of 22185 bp in length
            *    93678    93777: gap of 100 bp
            *    93778   101598: contig of 7821 bp in length
            *   101599   101698: gap of 100 bp
            *   101699   128689: contig of 26991 bp in length
            *   128690   128789: gap of 100 bp
            *   128790   138632: contig of 9843 bp in length
            *   138633   138732: gap of 100 bp
            *   138733   144193: contig of 5461 bp in length
            *   144194   144293: gap of 100 bp
            *   144294   155818: contig of 11525 bp in length
            *   155819   155918: gap of 100 bp
            *   155919   168354: contig of 12436 bp in length
            *   168355   168454: gap of 100 bp
            *   168455   171301: contig of 2847 bp in length
            *   171302   171401: gap of 100 bp
            *   171402   180364: contig of 8963 bp in length
            *   180365   180464: gap of 100 bp
            *   180465   188604: contig of 8140 bp in length
            *   188605   188704: gap of 100 bp
            *   188705   192855: contig of 4151 bp in length
            *   192856   192955: gap of 100 bp
            *   192956   201754: contig of 8799 bp in length
            *   201755   201854: gap of 100 bp
            *   201855   211372: contig of 9518 bp in length
            *   211373   211472: gap of 100 bp
            *   211473   225581: contig of 14109 bp in length.
FEATURES             Location/Qualifiers
     source          1. .225581
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-3L24"
                     /clone_lib="DanioKey"
     misc_feature    1. .20602
                     /note="assembly_fragment:01463
                     fragment_chain:1"
     misc_feature    20703. .57831
                     /note="assembly_fragment:01892
                     fragment_chain:1"
     misc_feature    57932. .65164
                     /note="assembly_fragment:02319
                     fragment_chain:1"
     misc_feature    65265. .71392
                     /note="assembly_fragment:02166
                     fragment_chain:1"
     misc_feature    71493. .93677
                     /note="assembly_fragment:01041
                     fragment_chain:1"
     misc_feature    93778. .101598
                     /note="assembly_fragment:02235
                     fragment_chain:1"
     misc_feature    101699. .128689
                     /note="assembly_fragment:01236
                     fragment_chain:1"
     misc_feature    128790. .138632
                     /note="assembly_fragment:02758
                     fragment_chain:1"
     misc_feature    138733. .144193
                     /note="assembly_fragment:02245
                     fragment_chain:1"
     misc_feature    144294. .155818
                     /note="assembly_fragment:01852
                     fragment_chain:1"
     misc_feature    155919. .168354
                     /note="assembly_fragment:01722
                     fragment_chain:1"
     misc_feature    168455. .171301
                     /note="assembly_fragment:00772
                     fragment_chain:2"
     misc_feature    171402. .180364
                     /note="assembly_fragment:02453
                     fragment_chain:2"
     misc_feature    180465. .188604
                     /note="assembly_fragment:00524
                     fragment_chain:2"
     misc_feature    188705. .192855
                     /note="assembly_fragment:02841
                     fragment_chain:2"
     misc_feature    192956. .201754
                     /note="assembly_fragment:00644
                     fragment_chain:2"
     misc_feature    201855. .211372
                     /note="assembly_fragment:02536
                     fragment_chain:2"
     misc_feature    211473. .225581
                     /note="assembly_fragment:01373
                     fragment_chain:2"
ORIGIN

  Query Match             6.8%;  Score 75.6;  DB 14;  Length 225581;
  Best Local Similarity   45.1%;  Pred. No. 0.0021;
  Matches  279;  Conservative    0;  Mismatches  339;  Indels    0;  Gaps    0;

Qy         34 ATAAGAGAAAAAGCATTTCCTTTTAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAAGG 93
```

```
Db     151603 AGACAAGTAAAAAAAATTCTATTTAAACATTAAAAAACATATATTTTTAATATTTCTAAC 151662

Qy         94 CCTCGATGGATTTCATTAAATAGATTTAATGAGCAATACTATAGAGATGCCTTTACACAA 153

Db     151663 CCCCCTTTAAAACCATCCGATGCTTGTATAGTGCAAATGTTCTGGGAGATCAACACTACT 151722

Qy        154 AATAATATAGAGACGAATCTTACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATAT 213

Db     151723 ACTTCTACTACTACTACTACTTATAATAATAATAATAATAATAATAATAATAATAATGAT 151782

Qy        214 TTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTG 273

Db     151783 AATAATAATAATGATAATAATAATAATAATAATGATAAAAATGATAATAATAATAATGAT 151842

Qy        274 AATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAACACTGTTACTATCCTTCATTT 333

Db     151843 AATAATAATAATGATAATAATAATAATAATAATGATAATAATAATAATAATAATAATAAT 151902

Qy        334 TTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCT 393

Db     151903 AATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGTTATTAATAATAAT 151962

Qy        394 AACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAAT 453

Db     151963 ATTGATAATAATGATAATAATAATAATGATAATAATAATAATAATGATAATAATAATAAT 152022

Qy        454 TATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCT 513

Db     152023 AATAATAATAATAATAATGATAATAATAATAATAATGATAATAATAATAATAATAATAAT 152082

Qy        514 GATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCAAT 573

Db     152083 AATGATAATAATAATGATAATAATAATAATAATAATGATAATAATGATAATAATAATAAT 152142

Qy        574 CAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAA 633

Db     152143 AATGATAATAATAATAATAATAATAATAATGATAATAATAATAATAATAATAATAATGAT 152202

Qy        634 TATAATCCTGATGTTTAT 651

Db     152203 AATAATAATGATAATAAT 152220
```

```
RESULT 14
BX890614
LOCUS       BX890614              115990 bp    DNA     linear   VRT 13-AUG-2004
DEFINITION  Zebrafish DNA sequence from clone DKEY-46N5, complete sequence.
ACCESSION   BX890614
VERSION     BX890614.10  GI:51241707
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 115990)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 13, 2004 this sequence version replaced gi:47776053.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-46N5
            is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5.
FEATURES             Location/Qualifiers
     source          1. .115990
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-46N5"
                     /clone_lib="DanioKey"
ORIGIN
```

```
  Query Match             6.6%;  Score 73.6;  DB 5;  Length 115990;
  Best Local Similarity   44.8%;  Pred. No. 0.0055;
  Matches  368;  Conservative    0;  Mismatches  449;  Indels    5;  Gaps    2;

Qy          2 TTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAAA 61

Db      54864 TTCTTTAAAGTATAACTAATATAAATAATATAATAATAATAATAATAATAATAATAATAA 54923

Qy         62 GTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTTA 121

Db      54924 TAATAATAATAATAATGATAATAATAATAATGATAATAATAATAATAAT-AATAATGATA 54982

Qy        122 ATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTTA 181

Db      54983 ATAATAATAATAATAATAATAATAATGATAATAATAATAATGATGATAATAATAATAATA 55042

Qy        182 TTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATTC 241

Db      55043 ATAATAATGATGATGATAATAATAAAGAAAATAATGATAATAATAATGATAATAATAATA 55102

Qy        242 TTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTT 301

Db      55103 ATAATAATAATAATAATAATGATAATAATGATAATAATAATAATAATAATAATAATGATG 55162

Qy        302 ATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGA 361

Db      55163 ATGATAATAATAATAATAATAATAATGATAATAATAATAATGATGATGATAATAATAATA 55222

Qy        362 ATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAG 421

Db      55223 ATAATAATAATAATAATAATAATAATAATAATAATGATGATAATAATAATAATAATAATA 55282

Qy        422 TTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCAT 481

Db      55283 AAAATAATGATAATAATAATAATAATAATAATAATGATAATAATAATAATGATGATGATA 55342

Qy        482 TTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATA 541

Db      55343 ATAATAATAATAATAATAATGATAATAATGATAATAATAATAATAATAATAATAATAATG 55402
```

```
Qy        542 ATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAA 601
              ||  ||| ||  || ||  ||  | |   ||| |   |   ||  || | ||| ||   |
Db      55403 ATGATGATAATAATAATAATAATAATAATAATGATAATAATAATAATGATGATGATAATA 55462

Qy        602 CTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGCA 661
              |  |  ||       ||| | | ||| |||  ||||  | ||  | ||         |
Db      55463 ATAATAATAATAATAATAATAATAATAATGATAATAATAATAATAATAATAATAAAA 55522

Qy        662 CGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGA----ATTGATGGTTTTG 717
              | || | | ||| ||         |    ||  |||| |    | ||||| |  |
Db      55523 ATAATGATAATAATGATAATAATAATAATAATGATAATAATAATAATAATGATGATGATA 55582

Qy        718 GCAAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAA 777
              || ||   |  |||   |  |  || |  ||| |       ||| | |||       |
Db      55583 ATAATAATAATAATAATAATAATAATGATAATAATAATGATAATAATAACAAAAACAACA 55642

Qy        778 CAAATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCA 819
              || | |  || |  ||  |||||  || |   ||   || ||
Db      55643 CATAATCCATCAGACATTAATTGTTTTTCAAAAACCAAAACA 55684


RESULT 15
CR848672
LOCUS       CR848672              155885 bp    DNA     linear   HTG 14-MAY-2005
DEFINITION  Danio rerio clone CH211-166E11, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   CR848672
VERSION     CR848672.3  GI:65301290
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 155885)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 14, 2005 this sequence version replaced gi:54260976.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC166E11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 153972 bases at least Q40
            Consensus quality: 154337 bases at least Q30
            Consensus quality: 154522 bases at least Q20
            Insert size: 155385; sum-of-contigs
            Insert size: 153983; 2.8% error; agarose-fp
            Quality coverage: 7.38x in Q20 bases; sum-of-contigs Quality
            coverage: 7.70x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10134: contig of 10134 bp in length
            *    10135    10234: gap of 100 bp
            *    10235   111588: contig of 101354 bp in length
            *   111589   111688: gap of 100 bp
            *   111689   123934: contig of 12246 bp in length
            *   123935   124034: gap of 100 bp
            *   124035   133984: contig of 9950 bp in length
            *   133985   134084: gap of 100 bp
            *   134085   149579: contig of 15495 bp in length
            *   149580   149679: gap of 100 bp
            *   149680   155885: contig of 6206 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155885
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-166E11"
                     /clone_lib="CHORI-211"
     misc_feature    1. .10134
                     /note="assembly_fragment:00183
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    10235. .111588
                     /note="assembly_fragment:00620
                     fragment_chain:1"
     misc_feature    111689. .123934
                     /note="assembly_fragment:00082
                     fragment_chain:1"
     misc_feature    124035. .133984
                     /note="assembly_fragment:00296
                     fragment_chain:1"
     misc_feature    134085. .149579
                     /note="assembly_fragment:00428
                     fragment_chain:1"
     misc_feature    149680. .155885
                     /note="assembly_fragment:00029
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             6.6%;  Score 73.6;  DB 14;  Length 155885;
  Best Local Similarity   46.0%;  Pred. No. 0.005;
  Matches  286;  Conservative    0;  Mismatches  334;  Indels    2;  Gaps    1;

Qy         32 TTATAAGAGAAAAAGCATTTCCTTTTAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAA 91
              | |||| |  || |  | |      | | ||  ||| | | |  |||  ||| || |  ||
Db      46482 TAATAATAATAATAATAATAATAATAATGATAATAATAATGATAATAATAATAATGATGATAA 46541

Qy         92 GGCCTCGATGGATTTCATTAAATAGATT--TAATGAGCAATACTATAGAGATGCCTTTAC 149
                 |   || |   |||  | |||  ||||||  |  |  |||   ||      ||
Db      46542 TAATAATAATAATATTGATAATAATATTAATAATGATGATGATAATAATAATAATGATAA 46601

Qy        150 ACAAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTC 209
              |   |||||||   | |  |||  ||     || |||   ||| || | |    | ||
Db      46602 AAGTAATAATAATAATAATAATAATAATAATAATAATTATAATAGTTATAATAATAATAA 46661

Qy        210 ATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAA 269
              ||  | ||| ||   ||  |||||  |  |  |  |    |  |  |  |       |
Db      46662 TAATAATAATAATAATAATAATGATGATGATAATAATGATAATAATAATAATAATAATGA 46721

Qy        270 TTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAACACTGTTACTATCCTTC 329
              |   |||||  |       | |  | ||| | || |   | |  |   | |  ||  |
Db      46722 TAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGATAA 46781

Qy        330 ATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTG 389
              | ||| |||     ||      || | |  |||    |  |||    |     |||
Db      46782 TAATATTAATAATAATGATGATAATAATAATAATGATAATAGTAATAATAATAATAATAA 46841

Qy        390 CGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGG 449
              ||| |||| || |  ||||    ||  | || ||   ||  ||| |  ||  |
Db      46842 TAATAATAATAATAATAATAATAGTAATAATGCTAATAATAATAATAATGATAATAATAA 46901

Qy        450 GAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTT 509
              | ||||  |   ||     ||  |  |      |  |       | |    ||
Db      46902 TAGTTATAATAATAATAATAATAATAATGATGATGATAATAATGATAATAATAATAATAA 46961
```

```
Qy        510 CGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTG 569
              | | || |   |  ||||| ||  ||  |||||| | | ||  || ||  ||  | |
Db      46962 TGATAATGATAATAATAATAATGATAATAATAATAATAATAATAATAATAATAATAA 47021

Qy        570 CAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATAC 629
              || |   |     |  |||| ||| |||  | || |  ||       ||| | | |||
Db      47022 TGATAATAATAATGATAATAATGATAATGATAATGATAATAATAATAATAATAATAA 47081

Qy        630 TGAATATAATCCTGATGTTTAT 651
              | |  |||||  | || || ||
Db      47082 TAATAATAATAATAATATTAAT 47103


Search completed: May 10, 2006, 20:33:47
Job time : 4024.18 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:42:47 ; Search time 516.763 Seconds
(without alignments)
14276.972 Million cell updates/sec

Title: US-09-423-093B-2_COPY_5365_6471
Perfect score: 1107
Sequence: 1 attcttttaccgagaaatga.....aaattgaatcccgatcttaa 1107

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1107	100.0	14024	2	AAX06749	Aax06749 E. coli O
	2	1107	100.0	14024	3	AAZ56386	Aaz56386 Escherich
c	3	1107	100.0	14181	10	ADC01214	Adc01214 Enterohae
c	4	1107	100.0	14187	9	ACD19108	Acd19108 E. coli 0
c	5	64.8	5.9	4985	6	ABQ75107	Abq75107 Anopheles
c	6	64.8	5.9	4985	10	ACF79720	Acf79720 Mosquito
c	7	60	5.4	2000	11	ACL35887	Acl35887 Rice stre
c	8	59.2	5.3	8056	8	ABZ10246	Abz10246 Haematopo
c	9	56.6	5.1	8056	8	ABZ10100	Abz10100 Haematopo
	10	56.2	5.1	16217	6	ABL32625	Abl32625 Human imm
	11	56.2	5.1	335913	5	AAI61371	Aai61371 Soybean 2
	12	56.2	5.1	335913	5	AAI61372	Aai61372 Soybean 2
	13	55.4	5.0	11790	6	ABL32543	Abl32543 Human imm
	14	54.2	4.9	5487	6	ABL33598	Abl33598 Human imm
	15	54	4.9	6132	6	ABL32863	Abl32863 Human imm
c	16	53.4	4.8	11143	4	ABL12834	Abl12834 Drosophil
	17	52.6	4.8	2000	11	ACL37108	Acl37108 Rice stre
c	18	52.2	4.7	2000	11	ACL37108	Acl37108 Rice stre
	19	52	4.7	50000	6	ABL55643	Abl55643 AmEPV gen
	20	51.4	4.6	34688	6	ABQ67060	Abq67060 Human ang
c	21	51.2	4.6	1428	14	ADZ70899	Adz70899 Human mat
	22	51.2	4.6	2000	11	ACL35887	Acl35887 Rice stre
	23	51.2	4.6	8056	8	ABZ10246	Abz10246 Haematopo
c	24	51	4.6	1086	14	ADZ71081	Adz71081 Human chr
c	25	51	4.6	5145	3	AAA70209	Aaa70209 Plasmodiu
	26	50.8	4.6	21537	6	ABL33999	Abl33999 Human imm
	27	50.8	4.6	110000	6	ABA92787_3	Continuation (4 of
	28	50.6	4.6	7025	6	ABK40059	Abk40059 Human che
	29	50.6	4.6	7025	6	AAS63350	Aas63350 Chemicall
	30	50.2	4.5	1428	14	ADZ70899	Adz70899 Human mat
	31	49.6	4.5	2482	14	ADZ71091	Adz71091 Human chr
	32	49.4	4.5	1410	6	ABL56209	Abl56209 AmEPV DNA
	33	49.4	4.5	8961	6	ABK28428	Abk28428 DNA trans
	34	49.4	4.5	8961	6	ABL49380	Abl49380 Human pol
	35	49.4	4.5	50000	6	ABL55644	Abl55644 AmEPV gen
	36	49.4	4.5	70768	6	AAL41152	Aal41152 Wooden le
	37	49	4.4	2000	11	ACL35363	Acl35363 Rice stre
	38	49	4.4	11143	4	ABL12834	Abl12834 Drosophil
	39	48.8	4.4	883	4	AAL15210	Aal15210 Human bre
	40	48.8	4.4	960	11	ACN85231	Acn85231 Breast ca
c	41	48.8	4.4	1508	14	ADZ70993	Adz70993 Human chr
	42	48.8	4.4	61020	4	AAS46788	Aas46788 Tumour su
	43	48.6	4.4	5398	6	ABL32142	Abl32142 Human imm
	44	48.4	4.4	2310	4	AAS46734	Aas46734 Tumour su
	45	48.4	4.4	4001	13	ADS89759	Ads89759 Oligonucl

ALIGNMENTS

RESULT 1
AAX06749
ID AAX06749 standard; DNA; 14024 BP.
XX
AC AAX06749;
XX
DT 26-APR-1999 (first entry)
XX
DE E. coli O157 antigen gene cluster.
XX
KW O antigen; O157 antigen; diagnosis; wbdN gene; wzx gene; wzy gene;
KW wbdP gene; wbdO gene; wbdR gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 79. .861
FT /*tag= a
FT /product= "sugar_transferase"
FT /note= "wbdN gene (ORF1), this region is specifically
FT claimed in Claim 6; encodes AAW88312"
FT CDS 858. .2042
FT /*tag= b
FT /product= "O_antigen-polymerase"
FT /note= "wzy gene (ORF2), this region is specifically
FT claimed in Claim 6; encodes AAW88313"
FT CDS 2011. .2757
FT /*tag= c
FT /product= "sugar_transferase"
FT /note= "wbdO gene (ORF3), this region is specifically
FT claimed in Claim 6; encodes AAW88314"
FT CDS 2744. .4135
FT /*tag= d
FT /product= "O_antigen_flippase"
FT /note= "wzx gene (ORF4), this region is specifically
FT claimed in Claim 6; encodes AAW88315"
FT CDS 4132. .5232
FT /*tag= e
FT /product= "perosamine_synthetase"
FT /note= "per gene (ORF5); encodes AAW88316"
FT CDS 5257. .6471

```
FT                   /*tag=  f
FT                   /product= "sugar_transferase"
FT                   /note= "wbdP gene (ORF6), this region is specifically
FT                   claimed in Claim 6; encodes AAW88317"
FT   CDS             6491. .7609
FT                   /*tag=  g
FT                   /note= "ORF7; encodes AAW88318"
FT   CDS             7606. .8578
FT                   /*tag=  h
FT                   /note= "ORF8; encodes AAW88319"
FT   CDS             8580. .9089
FT                   /*tag=  i
FT                   /note= "ORF9; encodes AAW88320"
FT   CDS             9071. .10519
FT                   /*tag=  j
FT                   /note= "manC gene (ORF10); encodes AAW88321"
FT   CDS             10523. .11893
FT                   /*tag=  k
FT                   /note= "manD gene (ORF11); encodes AAW88322"
FT   repeat_region   12007. .13123
FT                   /tag= l
FT                   /note= "H-repeat, may play a role in gene cluster
FT                   assembly"
FT   CDS             13156. .13821
FT                   /*tag=  m
FT                   /product= "N-acetyl_transferase"
FT                   /note= "wbdR gene (ORF12), this region is specifically
FT                   claimed in Claim 6; encodes AAW88323"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88312, AAW88313, AAW88314, AAW88315, AAW88316, AAW88317,
DR   AAW88318, AAW88319, AAW88320, AAW88321, AAW88322, AAW88323.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 6; Fig 8; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Escherichia coli O157 O antigen. The gene cluster was
CC   obtained by PCR amplification of E. coli O157:H7 strain (C664-1992) DNA
CC   using primers (see AAX06752-63) based on O antigen gene sequences. 12
CC   Open reading frames within the gene cluster were identified on the basis
CC   of homology to known sequences. The genes encode proteins (see AAW88312-
CC   23) involved the synthesis of sugars present in the polysaccharide
CC   antigen, and in the transport or processing of polysaccharide or
CC   oligosaccharide units. The use of nucleic acid molecules derived from
CC   particular assembly and transport genes, particularly wbd (transferase),
CC   wzx (flippase) and wzy (polymerase) genes, within O antigen gene clusters
CC   improves the specificity of methods for the detection and identification
CC   of O antigens, e.g. in tests of food- or faecal- derived samples, or
CC   samples from patients. The O antigen is a major virulence factor of
CC   enteropathogenic E. coli strains that cause diarrhoea and haemorrhagic
CC   colitis
XX
SQ   Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 1107;  DB 2;  Length 14024;
  Best Local Similarity  100.0%;  Pred. No. 6.4e-222;
```

```
  Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5365 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 5424

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5425 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 5484

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5485 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 5544

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5545 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 5604

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5605 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 5664

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5665 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 5724

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5725 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 5784

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5785 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 5844

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5845 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 5904

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5905 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 5964

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5965 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 6024

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6025 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 6084

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6085 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 6144

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6145 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 6204

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6205 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 6264

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6265 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 6324

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6325 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 6384

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6385 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 6444
```

```
Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
             ||||||||||||||||||||||||||||
Db       6445 AAACAAGAAATTGAATCCCGATCTTAA 6471
```

RESULT 2
AAZ56386
ID AAZ56386 standard; DNA; 14024 BP.
XX
AC AAZ56386;
XX
DT 17-MAR-2000 (first entry)
XX
DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:56.
XX
KW Flagellin; fliC; antigen; detection; ds.
XX
OS Escherichia coli.
XX
PN WO9961458-A1.
XX
PD 02-DEC-1999.
XX
PF 21-MAY-1999; 99WO-AU000385.
XX
PR 21-MAY-1998; 98AU-00003634.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Reeves PR, Wang L;
XX
DR WPI; 2000-072598/06.
XX
PT Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, feces, etc.
XX
PS Claim 3; Page 229-233; 245pp; English.
XX
CC AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all or
CC part of an Escherichia coli flagellin protein except a protein expressed
CC by E. coli H1, H7, H12 or H48 type strains. The present invention also
CC describes a method of detecting the presence of E. coli of a particular H
CC serotype in a sample, comprising specifically hybridising a nucleic acid,
CC preferably at least a pair, derived from a flagellating gene, specific
CC for a particular flagellin gene associated with the H serotype, to any
CC E.coli in the sample which contain the gene, and detecting any hybridised
CC molecules, identifying the presence of that serotype in the sample. (I)
CC are useful for: (1) detecting the presence of E. coli of H serotype in a
CC sample by hybridising at least one or a pair of (I) to any E. coli in the
CC sample and detecting the hybridised nucleic acid molecules; and (2) for
CC detecting the presence of both O and H-serotypes of E. coli by
CC hybridising at least one or a pair of (I) to any E. coli present in the
CC sample and detecting the hybridised nucleic acid molecules. (I) is
CC particularly useful for detecting the combination of O and H antigen.
CC Hybridised (I) when using at least one (I) is detected by southern blot
CC analysis and, when using a pair of (I), is detected by polymerase chain
CC reaction (PCR). AAZ56399 to AAZ56420 represent primers used in the
CC exemplification of the present invention
XX
SQ Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 U; 0 Other;

  Query Match   100.0%; Score 1107; DB 3; Length 14024;
  Best Local Similarity 100.0%; Pred. No. 6.4e-222;
  Matches 1107; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5365 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 5424

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5425 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 5484

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5485 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 5544

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5545 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 5604

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5605 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 5664

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5665 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 5724

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5725 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 5784

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5785 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 5844

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5845 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 5904

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5905 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 5964

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5965 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 6024

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6025 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 6084

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6085 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 6144

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6145 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 6204

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6205 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 6264

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6265 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 6324

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6325 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 6384

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6385 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 6444

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
              ||||||||||||||||||||||||||||
Db       6445 AAACAAGAAATTGAATCCCGATCTTAA 6471
```

RESULT 3
ADC01214/c
ID ADC01214 standard; DNA; 14181 BP.

```
XX
AC   ADC01214;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Enterohaemorragic E. coli 0157:H7-specific nucleic acid SEQ ID NO: 1258.
XX
KW   ds; gene; enterohaemorragic; anti-bacterial.
XX
OS   Escherichia coli; 0157:H7.
XX
PN   JP2002355074-A.
XX
PD   10-DEC-2002.
XX
PF   24-JAN-2002; 2002JP-00015959.
XX
PR   24-JAN-2001; 2001JP-00112010.
XX
PA   (UYTS-) UNIV TSUKUBA.
XX
DR   WPI; 2003-451640/43.
XX
PT   Enterohemorragic Escherichia coli O157:H7-specific nucleic acid molecule
PT   and a polypeptide and its use, a polypeptide, a vector and a host cell.
XX
PS   Claim 2; SEQ ID NO 1258; 2067pp; Japanese.
XX
CC   The invention relates to a novel enterohaemorragic Escherichia coli
CC   O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
CC   has anti-bacterial activity. The polypeptide can be used in detection
CC   and/or treatment of O157:H7 infection. The nucleotide sequence of the
CC   genome of Enterohaemorragic E coli O157:H7 was determined. The present
CC   sequence represents an E. coli 0157:H7-specific nucleic acid of the
CC   invention.
XX
SQ   Sequence 14181 BP; 4450 A; 2856 C; 2281 G; 4594 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 1107;  DB 10;  Length 14181;
  Best Local Similarity  100.0%;  Pred. No. 6.4e-222;
  Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8631 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 8572

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8571 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 8512

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8511 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 8452

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8451 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 8392

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8391 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 8332

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8331 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 8272

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8271 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 8212

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8211 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 8152

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8151 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 8092

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8091 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 8032

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8031 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 7972

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7971 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 7912

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7911 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 7852

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7851 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 7792

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7791 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 7732

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7731 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 7672

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7671 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 7612

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7611 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 7552

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
              ||||||||||||||||||||||||||
Db       7551 AAACAAGAAATTGAATCCCGATCTTAA 7525


RESULT 4
ACD19108/c
ID   ACD19108 standard; DNA; 14187 BP.
XX
AC   ACD19108;
XX
DT   27-OCT-2003  (revised)
DT   21-AUG-2003  (first entry)
XX
DE   E. coli 0157 unique DNA sequence OZID_121.
XX
KW   OZID; ds; acute haemorrhagic colitis; haemolytic uraemic syndrome;
KW   food poisoning.
XX
OS   Escherichia coli; strain 0157:H7.
XX
PN   US2003023075-A1.
XX
PD   30-JAN-2003.
XX
PF   01-APR-2002; 2002US-00114170.
XX
PR   04-DEC-1998;   98US-0110955P.
PR   03-DEC-1999;   99US-00453702.
XX
PA   (BLAT/) BLATTNER F R.
PA   (BURL/) BURLAND V D.
```

```
PA   (PERN/) PERNA N T.
PA   (PLUN/) PLUNKETT G.
PA   (WELC/) WELCH R.
XX
PI   Blattner FR,  Burland VD,  Perna NT,  Plunkett G,  Welch R;
XX
DR   WPI; 2003-479497/45.
XX
PT   New DNA sequences from Escherichia coli strain O157:H7, useful for
PT   detecting E. coli O157:H7 in a sample, or in designing diagnostic probes
PT   which can be used to distinguish strain O157:H7 from strain K12 using
PT   molecular techniques.
XX
PS   Claim 16; SEQ ID NO 121; 33pp; English.
XX
CC   The invention relates to an isolated DNA molecule comprising an E. coli
CC   strain 0157:H7 sequence selected from a clostridial cytotoxin-like gene,
CC   a urease gene cluster, a RTX toxin-like gene cluster, a locus of
CC   enterocyte effacement and 2 genes from its associated lymphocytic phage
CC   933W (a putative serine/threonine kinase and a tail fibre gene). E. coli
CC   0157:H7 can cause food poisoning, specifically acute haemorrhagic colitis
CC   (which can develop into haemolytic uraemic syndrome). Also included are
CC   an isolated DNA molecule comprising a nucleotide sequence identical to at
CC   least 25 contiguous nucleotides contained in DNA sequences selected from
CC   ACD18988-ACD19242 (being 255 E.coli 0157 DNA sequences which are not
CC   found in E.coli K12), a recombinant DNA construction comprising the DNA
CC   above and a method for detecting E. coli 0157:H7 (ATCC 43895) in a sample
CC   (or distinguishing between 0157 and K12) using a probe derived from one
CC   of the 255 sequences. The DNA sequences are useful in detecting E. coli
CC   O157:H7 in a sample, for the early diagnosis of humans and livestock
CC   infected with O157:H7, and in designing diagnostic probes which can be
CC   used to distinguish strain O157:H7 from strain K12 using molecular
CC   techniques. The present sequence is one of the 255 E. coli 0157:H7 DNA
CC   sequence (termed OZID_1-OZID255). Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from the USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030023075 (Updated on 27-OCT-2003
CC   to standardise OS field)
XX
SQ   Sequence 14187 BP; 4451 A; 2859 C; 2282 G; 4595 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1107;  DB 9;  Length 14187;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-222;
  Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8636 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 8577

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8576 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 8517

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8516 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 8457

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8456 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 8397

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8396 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 8337

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8336 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 8277

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8276 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 8217

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8216 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 8157

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8156 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 8097

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8096 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 8037

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8036 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 7977

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7976 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 7917

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7916 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 7857

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7856 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 7797

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7796 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 7737

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7736 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 7677

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7676 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 7617

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7616 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 7557

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
              ||||||||||||||||||||||||||||
Db       7556 AAACAAGAAATTGAATCCCGATCTTAA 7530


RESULT 5
ABQ75107/c
ID   ABQ75107 standard; cDNA; 4985 BP.
XX
AC   ABQ75107;
XX
DT   01-NOV-2002  (first entry)
XX
DE   Anopheles gambiae odourant receptor 2 genomic DNA SEQ ID NO:10.
XX
KW   Anopheles gambiae; mosquito; olfactory gene; arrestin 1; pest control;
KW   odourant receptor; olfaction; gene; ds.
XX
OS   Anopheles gambiae.
XX
PN   WO200259274-A2.
XX
PD   01-AUG-2002.
XX
PF   28-JAN-2002; 2002WO-US002549.
XX
PR   26-JAN-2001; 2001US-0264649P.
PR   24-JAN-2002; 2002US-00056405.
XX
```

PA (UYVA-) UNIV VANDERBILT.
XX
PI Zwiebel LJ;
XX
DR WPI; 2002-627421/67.
DR P-PSDB; ABP52835.
XX
PT New mosquito olfaction polypeptides and polynucleotides, useful for
PT mosquito management, i.e. controlling the pest and disease vectors, or
PT for identifying pest control agents.
XX
PS Disclosure; Fig 4a; 96pp; English.
XX
CC The present invention describes a purified Anopheles gambiae olfaction
CC polypeptide comprising a 383, 394, 380, 411, 412, 391, 157 or 401 residue
CC amino acid sequence (see ABP52833 to ABP52840) (S1), a conservatively
CC modified amino acid sequence of them, or a sequence of (S1) with at least
CC 20 consecutive residues. Also described: (1) an isolated polynucleotide
CC comprising: (a) a nucleotide sequence encoding the purified Anopheles
CC gambiae olfaction polypeptide; or (b) a nucleotide sequence that
CC hybridises under stringent conditions to a hybridisation probe comprising
CC a 1964, 1239, 1142, 1236, 1194, 1176, 474 or 1206 nucleotide sequence
CC (see ABQ75102 to ABQ75105 and ABQ75110 to ABQ75113) (S2), or its
CC complement; and (2) a method for identifying an agent that binds to
CC mosquito olfaction molecules comprising: (a) providing an isolated
CC mosquito olfaction molecule; (b) contacting a test agent with the
CC isolated mosquito olfaction molecule; and (c) detecting specific binding
CC of the test agent to the isolated mosquito olfaction molecule, where the
CC presence of specific binding identifies the test agent as a mosquito
CC olfaction-binding compound. The mosquito olfaction molecules are useful
CC for mosquito management, i.e. controlling this pest and disease vector. A
CC method from the present invention of screening for substances that
CC modulate arrestin-odourant receptor interaction is useful for identifying
CC pest control agents. The present sequence represents Anopheles gambiae
CC odourant receptor 2 genomic DNA from the present invention. N.B. The
CC features given in figure 4a are tentative and do not directly encode SEQ
CC ID NO:6 (ABP52835)
XX
SQ Sequence 4985 BP; 1319 A; 1058 C; 990 G; 1618 T; 0 U; 0 Other;

Query Match 5.9%; Score 64.8; DB 6; Length 4985;
Best Local Similarity 46.1%; Pred. No. 0.00074;
Matches 286; Conservative 0; Mismatches 332; Indels 2; Gaps 2;

```
Qy        32 TTATAAGAGAAAAAGCATTTCCTTTTAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAA 91
             | |||| |  |  |  | |     | | ||  | | ||  |  |||  |||  |    ||
Db      2980 TAATAACAATAGCAATAATAAGAATAATAATAACAATAGCAATAATAATAATAATAATAA 2921

Qy        92 GGCCTCGATGGATTTCATTAAATAGATTTAATGAGCAATACTATAGAGATGCCTTTACAC 151
                 |   ||   | |||  | | | ||  |  |  | ||||   |       ||
Db      2920 TAATAATAATAATAATAATAATAATAATAAAAAAAAATAATTATAATAAGAATAGTAATA 2861

Qy       152 AAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCAT 211
             | |||||||   | |  ||   ||  ||| ||| | ||| |    |     | ||
Db      2860 ATAATAATAATAATATTAAAAATAATATTAATAATAATAATAATAATAATAATAATAAGA 2801

Qy       212 ATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATT 271
             ||  | ||| ||   ||  | ||  || |  |   |    |   || |      || |
Db      2800 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATCATAATAAAATTA 2741

Qy       272 TGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAACACTGTTACTATCCTTCAT 331
             | | ||  | |  |  |  | | ||   ||  | |   || | | | |  ||  |  ||
Db      2740 TTATTATGATAATAATGTATAATAATGAATAACAATAATAATAAT-TAATAATAATAAAT 2682

Qy       332 TTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCG 391
             |  ||| | |   | |||||     ||  | | |  || ||| | | ||  | |||
Db      2681 TAATTAATTAATAATAAATAATAATAATTATTATTGTTATAATAATAAT-AATAATA 2623

Qy       392 CTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGA 451
               ||| |||| || |  ||||    |||    || ||   ||   |||||  ||  |   |
Db      2622 ATAATAATAATAATAATAATCATAATTATAATAATAATAATAATAATAATAATAATAATA 2563

Qy       452 ATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCG 511
```

```
             || ||  |   ||     ||  |  |      |||        | |   ||  |
Db      2562 ATAATAATAATAATAATAATAATAATAATAAAAGTAATAATTATAATAATAATAGTAATA 2503

Qy       512 CTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCA 571
               | || |   |  ||||| |   ||  |||||| | | ||  || ||  ||  | |   |
Db      2502 ATAATAATAATAATAATAATATTAATAATAATAATAATAATAATAATAATAATAATAATA 2443

Qy       572 ATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTG 631
             || | | |   ||  ||||  || ||   | |  |  ||        ||| | | ||| |
Db      2442 ATAATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2383

Qy       632 AATATAATCCTGATGTTTAT 651
             |  |||||  | ||  | ||
Db      2382 ATAATAATAATAATAATAAT 2363
```

RESULT 6
ACF79720/c
ID ACF79720 standard; DNA; 4985 BP.
XX
AC ACF79720;
XX
DT 15-JAN-2004 (first entry)
XX
DE Mosquito odorant receptor 2 genomic DNA.
XX
KW Mosquito; odorant receptor 2; olfaction; insecticide; antimalarial; gene;
KW ss.
XX
OS Anopheles gambiae.
XX

FH	Key	Location/Qualifiers
FT	exon	1. .702
FT		/*tag= d
FT		/number= 1
FT	5'UTR	1. .485
FT		/*tag= a
FT	TATA_signal	295. .300
FT		/*tag= b
FT	CDS	486. .3561
FT		/*tag= c
FT		/product= "Odorant receptor 2"
FT		/note= "contains introns"
FT	intron	703. .798
FT		/*tag= e
FT		/number= 1
FT	exon	799. .882
FT		/*tag= f
FT		/number= 2
FT	intron	883. .978
FT		/*tag= g
FT		/number= 2
FT	exon	979. .1404
FT		/*tag= h
FT		/number= 3
FT	intron	1405. .1601
FT		/*tag= i
FT		/number= 3
FT	exon	1602. .1702
FT		/*tag= j
FT		/number= 4
FT	intron	1703. .1783
FT		/*tag= k
FT		/number= 4
FT	exon	1784. .1886
FT		/*tag= l
FT		/number= 5
FT	intron	1887. .1993
FT		/*tag= m
FT		/number= 5
FT	exon	1994. .2106
FT		/*tag= n
FT		/number= 6

```
FT   intron          2107. .3459
FT                   /*tag=  o
FT                   /number= 6
FT   exon            3460. .4985
FT                   /*tag=  p
FT   3'UTR           3562. .4985
FT                   /*tag=  q
XX
PN   WO2003076590-A2.
XX
PD   18-SEP-2003.
XX
PF   10-MAR-2003; 2003WO-US007174.
XX
PR   08-MAR-2002; 2002US-00094240.
XX
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Zwiebel LJ;
XX
DR   WPI; 2003-722331/68.
XX
PT   New mosquito arrestin 1 and 2 genes and polypeptides, useful for
PT   identifying mosquito olfaction molecule binding compounds which reduce
PT   the ability of mosquitoes to locate sources of bloodmeal, e.g. humans.
XX
PS   Disclosure; Fig 4a; 101pp; English.
XX
CC   The present sequence is that of genomic DNA encoding a novel mosquito
CC   olfaction molecule, odorant receptor 2. A cDNA sequence for odorant
CC   receptor 2 is given in ACF79719. The invention provides 9 novel mosquito
CC   polypeptides and the nucleic acids encoding them. These are odorant
CC   receptor molecules 1-7 and arrestins 1 and 2. The odorant receptors
CC   function in a ligand-induced signal transduction pathway for the
CC   activation of mosquito olfaction. Arrestin functions to inhibit the
CC   activated signal transduction cascade. Thus, the odorant receptors act as
CC   an 'on' switch, and arrestin as an 'off' switch for the odorant detection
CC   system of the mosquito. Methods are provided for identifying compounds
CC   that interfere with the operation of the mosquito olfactory system,
CC   particularly compounds that modulate arrestin 2 activity. These are
CC   useful for the control of mosquitoes, particularly by reducing their
CC   ability to locate sources of bloodmeal
XX
SQ   Sequence 4985 BP; 1319 A; 1058 C; 990 G; 1618 T; 0 U; 0 Other;

  Query Match             5.9%;  Score 64.8;  DB 10;  Length 4985;
  Best Local Similarity   46.1%;  Pred. No. 0.00074;
  Matches  286;  Conservative    0;  Mismatches  332;  Indels    2;  Gaps    2;

Qy         32 TTATAAGAGAAAAAGCATTTCCTTTTAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAA 91
              | |||| |  |  |  | |     | | ||  | | ||  |  |||  |||  |    ||
Db       2980 TAATAACAATAGCAATAATAAGAATAATAATAACAATAGCAATAATAATAATAATAATAA 2921

Qy         92 GGCCTCGATGGATTTCATTAAATAGATTTAATGAGCAATACTATAGAGATGCCTTTACAC 151
                 |   ||   | |||  | | | ||  |  |  | ||||   |       ||
Db       2920 TAATAATAATAATAATAATAATAATAATAAAAAAAAATAATTATAATAAGAATAGTAATA 2861

Qy        152 AAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCAT 211
              | |||||||   | |  ||   ||   ||| ||| | ||| |     |      | ||
Db       2860 ATAATAATAATAATATTAAAAATAATATTAATAATAATAATAATAATAATAATAATAAGA 2801

Qy        212 ATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATT 271
              ||  | ||| ||   ||  | ||  || | |  |  |    |    | || |     || |
Db       2800 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATCATAATAAAATTA 2741

Qy        272 TGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAACACTGTTACTATCCTTCAT 331
              | | ||  | |  |  |  | | ||   ||  | |   || | | | |  ||  |  ||
Db       2740 TTATTATGATAATAATGTATAATAATGAATAACAATAATAATAAT-TAATAATAATAAAT 2682

Qy        332 TTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCG 391
              |  ||| | |   | |||||     ||  | | |  || ||| | | ||  | |||
Db       2681 TAATTAATTAATAATAAATAATAATAATAATTATTATTGTTATAATAATAAT-AATAATA 2623

Qy        392 CTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGA 451
              |||| |||| || |  ||||    |||    || ||   ||   |||||  ||  |   |
Db       2622 ATAATAATAATAATAATAATCATAATTATAATAATAATAATAATAATAATAATAATA 2563

Qy        452 ATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCG 511
              || ||  |   ||     ||  |  |     |||       | |   ||  |
Db       2562 ATAATAATAATAATAATAATAATAATAATAAAAGTAATAATTATAATAATAATAGTAATA 2503

Qy        512 CTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCA 571
              | || |   |  |||||  |   ||  ||||||  | | ||  || ||  ||  | |   |
Db       2502 ATAATAATAATAATAATAATATTAATAATAATAATAATAATAATAATAATAATAATA 2443

Qy        572 ATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTG 631
              || | | |   ||  ||||  || ||   | |  |  ||       ||| | | ||| |
Db       2442 ATAATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2383

Qy        632 AATATAATCCTGATGTTTAT 651
              |  |||||  | ||  | ||
Db       2382 ATAATAATAATAATAATAAT 2363


RESULT 7
ACL35887/c
ID   ACL35887 standard; cDNA; 2000 BP.
XX
AC   ACL35887;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Rice stress-regulated promoter SEQ ID NO:14450.
XX
KW   ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW   agriculture.
XX
OS   Oryza sativa.
XX
PN   WO2003008540-A2.
XX
PD   30-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US019668.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   24-AUG-2001; 2001US-0314662P.
PR   26-SEP-2001; 2001US-0325277P.
PR   21-NOV-2001; 2001US-0332132P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Kreps J,  Briggs SP,  Cooper B,  Glazebrook J,  Goff SA,  Katagiri F;
PI   Moughamer T,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-248011/24.
XX
PT   New stress-responsive nucleic acid, useful for altering the
PT   responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT   stress, salt stress or osmotic stress.
XX
PS   Claim 48; SEQ ID NO 14450; 89pp; English.
XX
CC   The invention relates to novel abiotic stress responsive polynucleotides
CC   and polypeptides. Also disclosed are vectors, expression cassettes, host
CC   cells, and plants containing such polynucleotides. Also disclosed are
CC   methods for using the polynucleotides and polypeptides to alter the
CC   responsiveness of a plant to abiotic stress. The invention is useful in
CC   agriculture. The nucleic acid is useful for determining whether a test
CC   plant has been exposed to an abiotic stress condition. It is also useful
CC   for selecting an agent that alters abiotic stress regulated
CC   polynucleotide expression in a plant cell, and to identify a homolog or
CC   ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC   molecule and the polypeptide encoded by it are useful in altering the
CC   responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC   stress, osmotic stress or any of their combinations. The present sequence
```

```
CC   is used in the exemplification of the invention
XX
SQ   Sequence 2000 BP; 410 A; 329 C; 253 G; 416 T; 0 U; 592 Other;

  Query Match             5.4%;  Score 60;  DB 11;  Length 2000;
  Best Local Similarity   17.3%;  Pred. No. 0.0065;
  Matches  157;  Conservative  321;  Mismatches  402;  Indels   28;  Gaps    2;

Qy         59 AAAGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGAT 118
Db       1018 AAAATAGACATATAGTACAACGWTATCGGTTAGTACCACGAAKAATTTGACTRATATAAC 959

Qy        119 TTAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCT 178
Db        958 TYATCGTATAATAATWTRMYTMATMTMAMWWMYSRWMYRATRMTAWAMWTMWWCTCRWAR 899

Qy        179 TTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTA 238
Db        898 YMKWASSKRTGYTWRRWWTWYWCYSRTMKWAWWAKGTGYKWTTRRYTWKYKSYGATSTWW 839

Qy        239 TTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATA 298
Db        838 AWATRTARCWAWKWRCTGMGTGKTKATKTAWTATAWRKAWMAMARTGCWRMKKGWWYAWW 779

Qy        299 TTTATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAA 358
Db        778 WWWTWYWRGKATWAYMMWKYMRMTSRMMYRWARKTTTCYKGWATKATMMAKWSRSKSMSM 719

Qy        359 GGAATGTGTTTTTTAAATTGATGCTCA---------ATTGCGCTAACAATATTATTGTTA 409
Db        718 YGKMSGWYYSSTKWMWGATRMWGMKWGKGCRRCGGYMKWSMKSKWTCASAWGMTKYKYSY 659

Qy        410 ATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTAC 469
Db        658 RCRGCGGTWTTRWWMRAWMCMAKWTKTGSYYWCRGYSKYKTTWYWRAWWWSCWKKTGTGR 599

Qy        470 ATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACT--------- 520
Db        598 TKYMWGTSTYGTTKMYSAYWKRMWSKKRWGRTKYATGWKYWGYTGWYCAYWSAAMCWKMR 539

Qy        521 ----------CTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGC 570
Db        538 KGTKTRTGRKSRGCWRTKCKTYWRMYRWMGYRWRTGTSRGRGGMAATKMSTKYRWKRWWR 479

Qy        571 AATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACT 630
Db        478 MAWATKSWKWRGMMWTKWCTKYAYKRWYMRAKAYGKKTTWGCSTYGKCTTWAYTRASYGA 419

Qy        631 GAATATAATCCTGATGTTTATTTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCT 690
Db        418 KWCTRTGKTTKGTCRGMKYRMTTAAGTKAKYYKAKRKTTTKTMWKMRYGSKWRWKTTAGY 359

Qy        691 GGATATTTTAATGAATTGATGGTTTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAG 750
Db        358 TRAKWKTYKYTRTTSWWRSKWKKTWWRKKWWRTKGTYKCTGKWRTMAMWWKKTARRTSWW 299

Qy        751 ATATTAGGGCATATACCTAAACTTGAACAAATTGAATTAATCAAAAATTGCATTGCTGTA 810
Db        298 KTKKYTTMKGKRWKCAMRAGTKAKWTCTRWWYYMYWAKRTRTGMMCKRSKGMWWWCTGAA 239

Qy        811 ATACAACCAACCTTATTTGAAGGCGGGCCTGGAGGGGGGGTAACATTTGACGCTATTGCA 870
Db        238 YCACWAYAKATGWYCTKYGRCAMACWTTGTYKTWSARWWYWTVKWTGAYWSKCAWTSWYW 179

Qy        871 TTAGGGAAAAAAGTTATACTATCTGACATAGATGTCAATAAAGAAGTTAATTGCGGTGAT 930
Db        178 YWKMWWRMAWAGRTGATWRWMAATYRCACRKCWTKWMWMGGTGAYAAWAAATTRKRMKRM 119

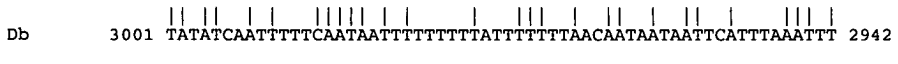
Qy        931 GTATATTT 938
Db        118 ATWTMTYW 111
```

```
RESULT 8
ABZ10246/c
ID   ABZ10246 standard; DNA; 8056 BP.
XX
AC   ABZ10246;
XX
DT   16-JAN-2003  (first entry)
XX
DE   Haematopoietic cell proliferation disorder related DNA sequence #386.
XX
KW   Human; haematopoietic cell proliferation disorder; cytostatic;
KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW   cytosine methylation state; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277272-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-EP003401.
XX
PR   26-MAR-2001; 2001US-0278333P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Berlin K,  Braun A,  Distler J,  Guetig D,  Howe A,  Mueller J;
PI   Olek A,  Piepenbrock C,  Adorjan P,  Grabs G,  Lesche R,  Leu E;
PI   Lewin A,  Lipscher E,  Maier S,  Model F,  Mueller V,  Otto T,  Pelet C;
PI   Schwope I,  Ziebarth H;
XX
DR   WPI; 2003-018942/01.
XX
PT   Detecting and differentiating between hematopoietic cell proliferative
PT   disorders, comprises contacting a target nucleic acid with a reagent that
PT   distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS   Claim 28; SEQ ID NO 386; 117pp; English.
XX
CC   The present invention describes a method for detecting and
CC   differentiating between haematopoietic cell proliferative disorders
CC   associated with at least 1 gene and/or their regulatory regions in a
CC   subject. The method comprises contacting a target nucleic acid in a
CC   biological sample obtained from the subject with at least 1 reagent,
CC   which distinguishes between methylated and non-methylated CpG
CC   dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC   represent specifically claimed nucleotide sequences from the present
CC   invention. Oligonucleotides from the present invention can be used: for
CC   differentiating between healthy haematopoietic cells and proliferative
CC   disorder haematopoietic cells; for differentiating between acute
CC   lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC   determining the cytosine methylation state and/or single nucleotide
CC   polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC   related sequences and their complements; and as primers for the
CC   amplification of haematopoietic cell proliferation disorder related DNA
CC   sequences. The nucleotide sequences from the present invention can also
CC   be used for detecting a predisposition to, differentiation between
CC   subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC   haematopoietic cell proliferative disorders. The present method enables a
CC   highly specific classification of haematopoietic cell proliferative
CC   disorders allowing for improved and informed treatment of patients
XX
SQ   Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;

  Query Match             5.3%;  Score 59.2;  DB 8;  Length 8056;
  Best Local Similarity   44.0%;  Pred. No. 0.012;
  Matches  342;  Conservative    0;  Mismatches  433;  Indels    3;  Gaps    2;

Qy         54 TTTTAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAA 113
Db       3061 TATAACATTTATTTTATTTTTAATTTATATTTATAAAATTTAAACATCATCATTTTAAAA 3002

Qy        114 TAGATTTAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCT 173
```



```
Db       3001 TATATCAATTTTTCAATAATTTTTTTTTATTTTTTTAACAATAATAATTCATTTAAATTT 2942

Qy        174 TACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGA 233

Db       2941 TACATTTCATAAAATTTTAACATTTTTTTTTCATTTTCAAATTTTTTATTTTTTTCAACA 2882

Qy        234 TGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGG 293

Db       2881 AAATTTTTTTCATAAAATAATTTCAAATTTTTAATATTTTATAAAAAAA--TATCATTTA 2824

Qy        294 TTATATTTATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGA 353

Db       2823 TTATTAATAACAATTTTAATAAACAAAATTCATTTAAATTTATTAAAAATAAACAATTAA 2764

Qy        354 TCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGC 413

Db       2763 ATATTTTTTTTATTTTTATTAATAATTTATTTAACAAATTATTTTTATTAAAAAAAAAAA 2704

Qy        414 ACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTC 473

Db       2703 ATAATAAATTTAAAAAAATTTAAAAATTTATTATTTTTAAAATTTTAAATTTTAACATTT 2644

Qy        474 TCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGC 533

Db       2643 TTTTTTATTTTTTTTTTTTAAATTATTTATATTTATAATTTTTTTTTTCATTTATTTTTT 2584

Qy        534 CAAATATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGA 593

Db       2583 ATTTTTTAAAAATTTTATTTAAATTTTTAAATTTTTTAATTTATTTTATAATTTAATTTA 2524

Qy        594 TCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTT 653

Db       2523 TTTTTTTTTATTTTTCATTAATATTTTAATTATTTTAAAACAAAATAATAAAAAATATAT 2464

Qy        654 AGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGT 713

Db       2463 A-AATAAAAATAAAAAAAAATTAATTTTTAAAAATAAATAAATTATTTTTTTTTTAATTT 2405

Qy        714 TTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACT 773

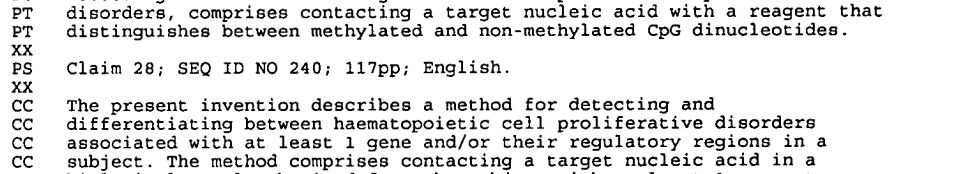
Db       2404 TCAAAAAATAATAAAATTTAATAAATTATTTATAAAAATAAAAAATTATATTTAAAAATT 2345

Qy        774 TGAACAAATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAA 831

Db       2344 AAATAAAATTTATTTAAATACAAAAATTAATAAATTATTTTAAAAATAAATAAATTAA 2287
```

RESULT 9
ABZ10100/c
ID ABZ10100 standard; DNA; 8056 BP.
XX
AC ABZ10100;
XX
DT 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #240.
XX
KW Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 240; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;

Query Match 5.1%; Score 56.6; DB 8; Length 8056;
Best Local Similarity 43.6%; Pred. No. 0.041;
Matches 459; Conservative 0; Mismatches 579; Indels 15; Gaps 4;



```
Qy          6 TTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAAAGTAT 65

Db       2466 TATAAATAAAAATAAAAAAAAAATTAATTTTTAAAAATAAATAAATTATTTTTTTTTTAAT 2407

Qy         66 ATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTTAATGA 125

Db       2406 TTTCGAAAAATAATAAAATTTAATAAATTATTTATAAAAATAAAAAATTATATTTA---- 2351

Qy        126 GCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTTATTAA 185

Db       2350 --AAAATTAAATAAAATTTATTTAAATACGAAAATTAATAAATTATTTTAAAAATAAATA 2293

Qy        186 AAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCC 245

Db       2292 AATTAAAAAATTAAAATTTTTTTAATTTATTTTAAATATTAAAAAATAAATTTTTATTAA 2233

Qy        246 TTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGA 305

Db       2232 AATTAAATTTTTTTTTTTTTTTATTTTTTTTTTCGTTTTTTTAAATTTTTAATAAAATTA 2173

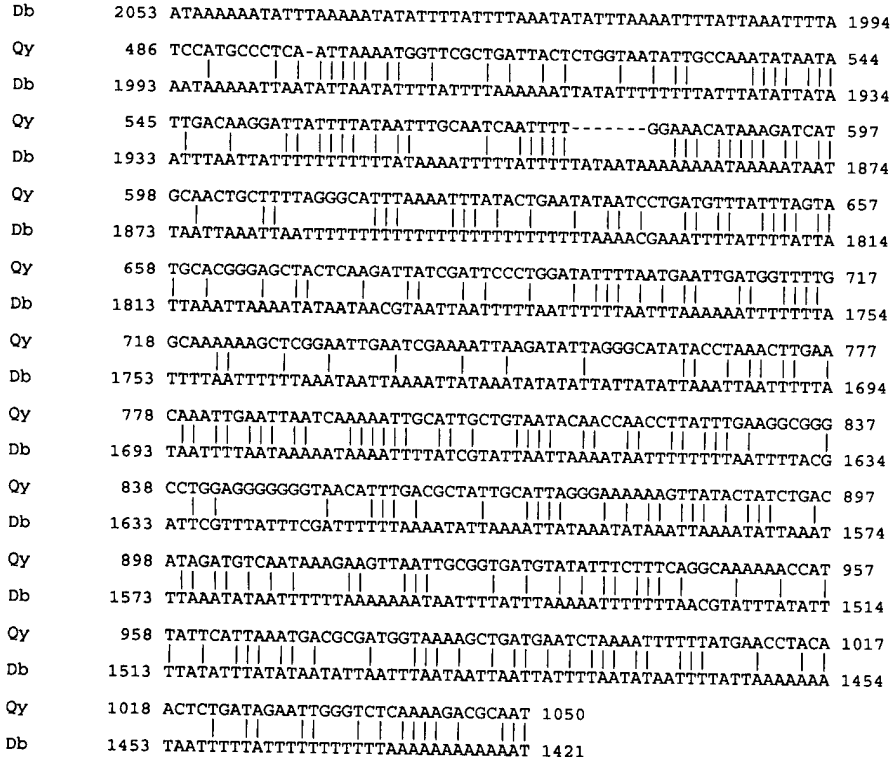
Qy        306 CTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGT 365

Db       2172 TAATTTTTTTTTTTTTTATTTTTTATTTTTAATAAAATAAAAAATTATTAAAATTATATT 2113

Qy        366 GTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTAT 425

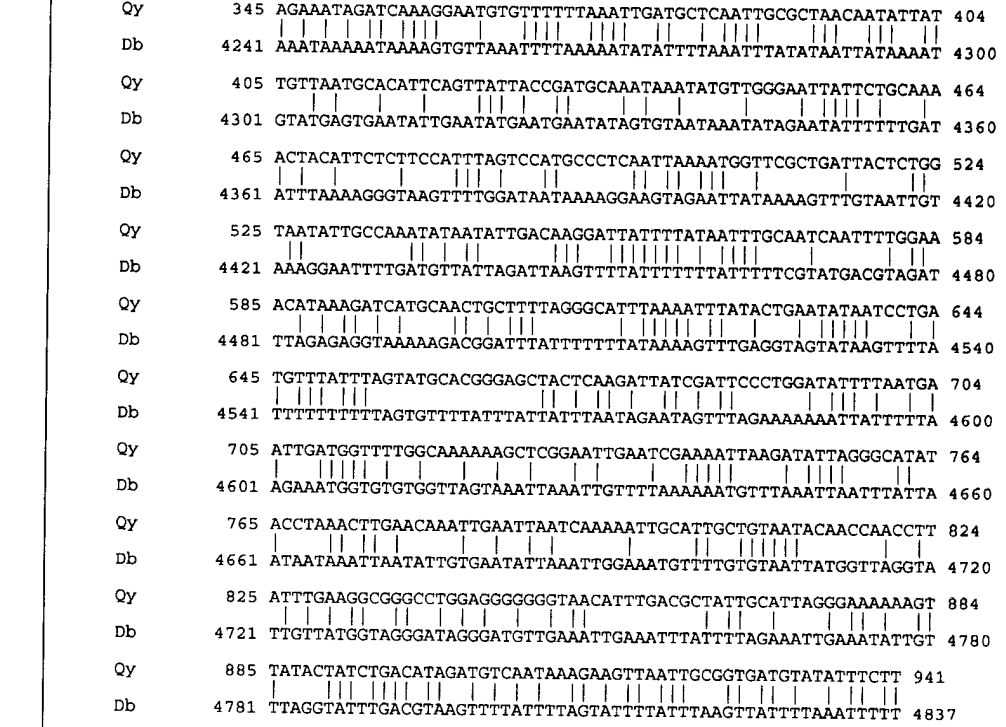
Db       2112 AAATAATTTA-TAATTATTATTTATAATATTTAAAATATTTTTAATAAAATTTTTTAAAA 2054

Qy        426 TACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAG 485
```



RESULT 10
ABL32625
ID ABL32625 standard; DNA; 16217 BP.
XX
AC ABL32625;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 598.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 598; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 16217 BP; 5200 A; 131 C; 2955 G; 7930 T; 0 U; 1 Other;

Query Match 5.1%; Score 56.2; DB 6; Length 16217;
Best Local Similarity 43.4%; Pred. No. 0.056;
Matches 259; Conservative 0; Mismatches 338; Indels 0; Gaps 0;



RESULT 11
AAI61371
ID AAI61371 standard; DNA; 335913 BP.
XX
AC AAI61371;

```
XX
DT   16-OCT-2001  (first entry)
XX
DE   Soybean 240017 region G3, SEQ ID NO: 2.
XX
KW   Soybean; antihelmintic; gene therapy; soybean cyst nematode; SCN;
KW   SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW   240017 region G3; 318013 region A3; 515002 region G2; ds.
XX
OS   Glycine max.
XX
PN   WO200151627-A2.
XX
PD   19-JUL-2001.
XX
PF   05-JAN-2001; 2001WO-US000552.
XX
PR   07-JAN-2000; 2000US-0174880P.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Hauge BM,  Wang ML,  Parsons JD,  Parnell LD;
XX
DR   WPI; 2001-425872/45.
DR   P-PSDB; AAM42214.
XX
PT   New purified nucleic acid for producing a soybean plant having soybean
PT   cyst nematode resistance and for use in plant breeding programs.
XX
PS   Claim 2; Page 204-400; 1353pp; English.
XX
CC   The invention relates to nucleic acid molecules from regions of the
CC   soybean genome which are associated with soybean cyst nematode (SCN)
CC   resistance. The nucleic acids are used to transform plants, and can
CC   produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC   The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC   of soybean plants and for introgressing SCN resistance or partial SCN
CC   resistance into soybean plants. They can also be used in plant breeding
CC   programmes. The invention also relates to proteins encoded by such
CC   nucleic acid molecules, as well as antibodies capable of recognising
CC   these proteins. The present sequence is a nucleic acid molecule provided
CC   in the specification
XX
SQ   Sequence 335913 BP; 114579 A; 53403 C; 53026 G; 114905 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 56.2;  DB 5;  Length 335913;
  Best Local Similarity   46.7%;  Pred. No. 0.089;
  Matches  178;  Conservative    0;  Mismatches  203;  Indels    0;  Gaps    0;

Qy        254 GTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAAC 313
              || ||   | |   ||||| |||| | || ||||  |   |||  || ||    ||  |
Db     223638 GTTTTATGTGGTTCAATTTTAATATAGAATCATGTGTGATTTAGTTTGATTCAATTTGAT 223697

Qy        314 ACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTA 373
                | ||       || | | | | || | | | || || ||  |   | |   | | |||
Db     223698 GATTTTTTATATTTTTAATATATTGTCAACTAAAACTAAATTTATATAGTTATTATGTTA 223757

Qy        374 AATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATG 433
              ||||  |    || ||   | |   | || | | ||       || |  ||||   | ||
Db     223758 AATTATTTTAAAAATGAATTTATTTTTTTTTATTAAAATTTGTTTAATATATTTTAGTTG 223817

Qy        434 CAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCC 493
              ||||  | ||| ||      ||||  |||| |  | |    |  |     ||   | |
Db     223818 AAAATTATTATTTTTACTTCTATTTGGCAATAACATAAAATAATATTACTAACTCTTTTT 223877

Qy        494 CTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGG 553
              | |  ||| |    |   ||  |  | || || |          | ||| || |   |
Db     223878 GTAACATAATAAAAATTTGTGTATCATTTGATATTTAATATTTTTTTAAAATAGTATATC 223937

Qy        554 ATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGG 613
              || ||| | |      | | || | |   |   || |  |     |||    |  ||
Db     223938 CTTTTTTAACATAAAAAAGTAAAATATATATTAATTATAAAATAACAAAAATTAATAAAT 223997

Qy        614 CATTTAAAATTTATACTGAAT 634
                ||| || | ||||  | |||
Db     223998 TATTCAATACTTATTATTAAT 224018


RESULT 12
AAI61372
ID   AAI61372 standard; DNA; 335913 BP.
XX
AC   AAI61372;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Soybean 240017 region G3, SEQ ID NO: 3.
XX
KW   Soybean; antihelmintic; gene therapy; soybean cyst nematode; SCN;
KW   SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW   240017 region G3; 318013 region A3; 515002 region G2; ds.
XX
OS   Glycine max.
XX
PN   WO200151627-A2.
XX
PD   19-JUL-2001.
XX
PF   05-JAN-2001; 2001WO-US000552.
XX
PR   07-JAN-2000; 2000US-0174880P.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Hauge BM,  Wang ML,  Parsons JD,  Parnell LD;
XX
DR   WPI; 2001-425872/45.
DR   P-PSDB; AAM42215.
XX
PT   New purified nucleic acid for producing a soybean plant having soybean
PT   cyst nematode resistance and for use in plant breeding programs.
XX
PS   Claim 2; Page 400-595; 1353pp; English.
XX
CC   The invention relates to nucleic acid molecules from regions of the
CC   soybean genome which are associated with soybean cyst nematode (SCN)
CC   resistance. The nucleic acids are used to transform plants, and can
CC   produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC   The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC   of soybean plants and for introgressing SCN resistance or partial SCN
CC   resistance into soybean plants. They can also be used in plant breeding
CC   programmes. The invention also relates to proteins encoded by such
CC   nucleic acid molecules, as well as antibodies capable of recognising
CC   these proteins. The present sequence is a nucleic acid molecule provided
CC   in the specification
XX
SQ   Sequence 335913 BP; 114582 A; 53398 C; 53027 G; 114906 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 56.2;  DB 5;  Length 335913;
  Best Local Similarity   46.7%;  Pred. No. 0.089;
  Matches  178;  Conservative    0;  Mismatches  203;  Indels    0;  Gaps    0;

Qy        254 GTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAAC 313
              || ||   | |   ||||| |||| | || ||||  |   |||  || ||    ||  |
Db     223638 GTTTTATGTGGTTCAATTTTAATATAGAATCATGTGTGATTTAGTTTGATTCAATTTGAT 223697

Qy        314 ACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTA 373
                | ||       || | | | | || | | | || || ||  |   | |   | | |||
Db     223698 GATTTTTTATATTTTTAATATATTGTCAACTAAAACTAAATTTATATAGTTATTATGTTA 223757

Qy        374 AATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATG 433
              ||||  |    || ||   | |   | || | | ||       || |  ||||   | ||
Db     223758 AATTATTTTAAAAATGAATTTATTTTTTTTTATTAAAATTTGTTTAATATATTTTAGTTG 223817

Qy        434 CAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCC 493
              ||||  | ||| ||      ||||  |||| |  | |    |  |     ||   | |
```

```
Db     223818 AAAATTATTATTTTTACTTCTATTTGGCAATAACATAAAATAATATTACTAACTCTTTTT 223877

Qy        494 CTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGG 553

Db     223878 GTAACATAATAAAAATTTGTGTATCATTTGATATTTAATATTTTTTTAAAATAGTATATC 223937

Qy        554 ATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGG 613

Db     223938 CTTTTTTAACATAAAAAAGTAAAATATATATTAATTATAAAATAACAAAAATTAATAAAT 223997

Qy        614 CATTTAAAATTTATACTGAAT 634

Db     223998 TATTCAATACTTATTATTAAT 224018


RESULT 13
ABL32543
ID   ABL32543 standard; DNA; 11790 BP.
XX
AC   ABL32543;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 516.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 516; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 11790 BP; 3627 A; 100 C; 2210 G; 5853 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 55.4;  DB 6;  Length 11790;
  Best Local Similarity   42.9%;  Pred. No. 0.078;
  Matches  325;  Conservative    0;  Mismatches  431;  Indels    1;  Gaps    1;

Qy         54 TTTTAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAA 113

Db       7019 TTTTAATATTTGGTAAATTAAATATATTTAGAGGGTTAGAAAATAGGGATTTTTTAGGAT 7078

Qy        114 TAGATTTAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCT 173

Db       7079 TTGTTAATAAATTTAATTAAAGGAAATTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAT 7138

Qy        174 TACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGA 233

Db       7139 TATTTTTAAGAGAGATAAATTTTTGAAAAATTAGTTTATTTTTTGAATTTTTGAATATAT 7198

Qy        234 TGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGG 293

Db       7199 GGGTTTTTTTTAATTGATTTGTTTAAATATTTATAGTTTTTATAGATAGTTAAAAGATAA 7258

Qy        294 TTATATTTATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGA 353

Db       7259 TTTTTATTTATTGTTTTTATATTTTTAGGTAGGTAATTTTTTTTTTTTACGTTTTTTTTTT 7318

Qy        354 TCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGC 413

Db       7319 TTTAATGATTAGATATTTGTTTTAAATGAGGTAATTAATTATTTTTTAAAATGTTATTTT 7378

Qy        414 ACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTC 473

Db       7379 TGTATTTAATTTAAGAGATATTAATAGAAATTTAGCGATTAGTTTTTTTAAATTGGATTT 7438

Qy        474 TCTTCCATTTAGTCCATGCCCTCAATT-AAAATGGTTCGCTGATTACTCTGGTAATATTG 532

Db       7439 TATTTTTTTTTTTATTTTTTTTTAATTGAAAATATTTGGAATTTTAAATATGTAAGTTTT 7498

Qy        533 CCAAATATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAG 592

Db       7499 TAAATTATATTTTAAAGAAAGTTATGTTTTAGTTTTAATTTAGATGTTTAGATGTTTAAG 7558

Qy        593 ATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATT 652

Db       7559 TAGATATTAATTTTTTAAAAATAGAGAAGAGGTAAATTGAACGATACGATTTTTTAGTTT 7618

Qy        653 TAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGG 712

Db       7619 ATTTTTTGAAATTAAAGATTTAAAATTTAGTAGGTAGAATAAGTTATAAATATAGAGTGT 7678

Qy        713 TTTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAAC 772

Db       7679 TTGAAGATTAAATGGGAGAAAAAGAATAAAGAGTATATACGTGATAGAATTAGTAGTTGT 7738

Qy        773 TTGAACAAATTGAATTAATCAAAAATTGCATTGCTGT 809

Db       7739 TTTTATAGTAAGAGTTTTTGAATTAGTGGGTTTTTTT 7775


RESULT 14
ABL33598
ID   ABL33598 standard; DNA; 5487 BP.
XX
AC   ABL33598;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1571.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
```

PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1571; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 5487 BP; 1532 A; 133 C; 1107 G; 2715 T; 0 U; 0 Other;

Query Match 4.9%; Score 54.2; DB 6; Length 5487;
Best Local Similarity 45.0%; Pred. No. 0.12;
Matches 294; Conservative 0; Mismatches 348; Indels 11; Gaps 2;

```
Qy        13 AGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAAAGTATATTAAAA 72
             | | ||| || | ||| | || |      ||      ||  || ||| |      || |
Db      2511 ATATATGTTAAATATTGTTTTGTTTTTTAAGTTATTGTTTATTGTAATATATTTTATT 2570

Qy        73 GCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAA----TAGATTTAATGAGCA 128
               | ||  ||||   ||||     |  ||   |||  | | |    || | |||||
Db      2571 AAATTTAAAAAGTAAGAAATTTATAAATTTTTTTTTTTAATAGTTTTAAAGTTAATTTTAT 2630

Qy       129 ATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTTATTAAAAG 188
               ||||         |||  | | |||  || |       |  |||  | | ||| ||
Db      2631 TATTTATAATATATATTTTTAAAAGAATGTTAGAATAGTTGAATTTATATATTTTATAAA 2690

Qy       189 TAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTG 248
               |     |  | |||| ||  | |     || |||      | || |  || ||  ||
Db      2691 AATTTATTGAGTTTTTTTTTTTTGTACGTGTATTAGGAGAAATTATATAGTTTTTTATTT 2750

Qy       249 CATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTT 308
               | |  | ||          || || || | |    |  |    || | || |||      |
Db      2751 TAAGGTTTTTATAGTATATAAATTTAAGATATTTGTGTATTTTAGGTAAATTATATTTTA 2810

Qy       309 TCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATA-------GATCAAAGGA 361
               || | | | | ||   || ||||| |  || |  |      |        | | | |  |
Db      2811 ATAAAAATATAAATAAAATTAATTTTGTTATATGTTATGGGTTTTTATAGGTTAAGAAAA 2870

Qy       362 ATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAG 421
               ||  |||| |  || ||  |  |||    |||  |  ||||  |  |  | ||| |
Db      2871 TTGATTTTTGGATTTTTATAATATATTTTTTAAATAGGTTATTTATTTTAGATATTTAT 2930

Qy       422 TTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCAT 481
             || ||  |||    |||  ||   ||   | | ||  |      | ||  ||   ||   |
Db      2931 TTTTTTTCGAGTAGAATGGATTAATTATAAGTAATGATATTGGATTAGTTTTGTTTTTTT 2990

Qy       482 TTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATA 541
               | |||  | |   | | |  || | |||     || | | |  ||| || |   | | ||
Db      2991 TAAGTATAGGTTTTTATTGTAATTTGTTTTTAAATAATTTTAATAAAATCGTTGATTTTA 3050

Qy       542 ATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAA 601
```

```
               | | | | || | |  ||  |  |||   |  |||||  |     |  | |      |
Db      3051 AGAGTTAAAAAGTTGTTTAGAGTTTTTTTTAGGGATTTTTAATTTTTTTTGTTATATGGA 3110

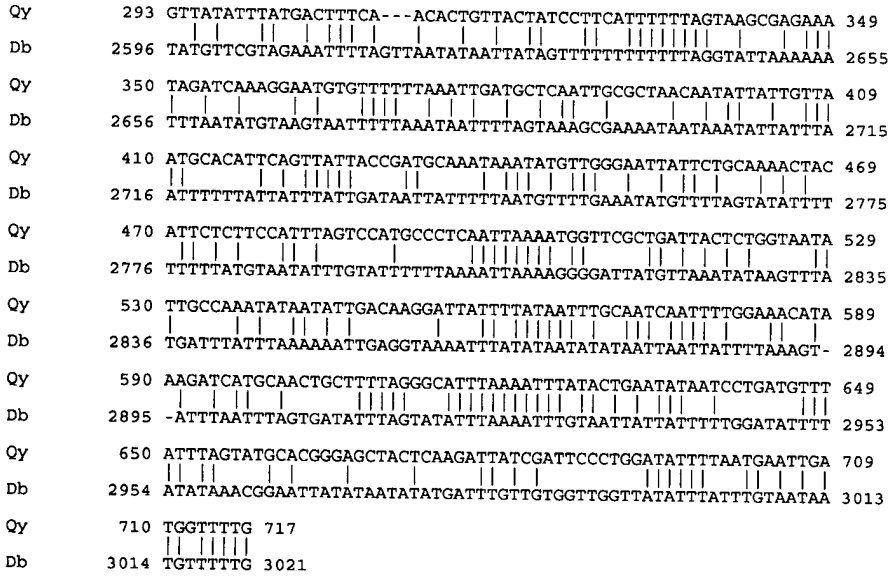
Qy       602 CTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTA 654
               |  ||  |       ||||| | |   |  | ||  |  |  ||   || ||
Db      3111 TTTTTTGAAATTTTGATAAAAGTAAAGTTATAGATTTTTTTTTTTGGAAATATA 3163
```

RESULT 15
ABL32863
ID ABL32863 standard; DNA; 6132 BP.
XX
AC ABL32863;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 836.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 836; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6132 BP; 1642 A; 28 C; 1246 G; 3216 T; 0 U; 0 Other;

Query Match 4.9%; Score 54; DB 6; Length 6132;
Best Local Similarity 46.2%; Pred. No. 0.14;
Matches 253; Conservative 0; Mismatches 290; Indels 5; Gaps 2;

```
Qy       173 TTACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTG 232
               ||   ||| |  | | | |  | | |       || |    |||  ||| | | |||| |
Db      2476 TTTGTTTTTTATAGATTGATTGTTATTATAAGTAGTGTGTTTTAAATATTTGGGATTATA 2535

Qy       233 ATGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTG 292
               |||   | |     |   | | |   || |   ||||| |   ||  ||| ||   ||
Db      2536 TTGTATATATAATGTAATTTATTTTATTTTTGATTAATTTTATAAATTAAGTATTTTTTT 2595
```



Search completed: May 10, 2006, 10:59:19
Job time : 520.763 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 11:00:02 ; Search time 3545.5 Seconds
(without alignments)
14608.170 Million cell updates/sec

Title: US-09-423-093B-2_COPY_5365_6471
Perfect score: 1107
Sequence: 1 attcttttaccgagaaatga.....aaattgaatcccgatcttaa 1107

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	76.6	6.9	1101	10	CNS0039G	AL063921 Drosophil
c	2	71	6.4	891	10	CW952279	CW952279 TcB38.1_B
c	3	68	6.1	1101	10	CNS00EVL	AL069706 Drosophil
c	4	65.4	5.9	1101	10	CNS0039G	AL063921 Drosophil
	5	64.4	5.8	471	10	CG807194	CG807194 1118078H0
	6	63.6	5.7	844	10	BX139987	BX139987 Danio rer
	7	62.6	5.7	889	10	CW973509	CW973509 AIAA-aaa5
c	8	62	5.6	641	9	AQ946120	AQ946120 Sheared D
	9	62	5.6	735	10	CW960936	CW960936 AIAA-aac1
c	10	60.8	5.5	1083	1	AJ926215	AJ926215 AJ926215
	11	60.6	5.5	1101	10	CNS00EVL	AL069706 Drosophil
	12	60.2	5.4	1027	10	CNS02T50	AL212733 Tetraodon
	13	60	5.4	969	10	CL469446	CL469446 SAIL_1303
c	14	60	5.4	1277	9	CC253231	CC253231 CH261-180
c	15	59.8	5.4	1101	10	CNS00EO7	AL069440 Drosophil
	16	59.8	5.4	1896	10	CG753083	CG753083 P048-1-C0
c	17	59.6	5.4	1380	1	AJ928744	AJ928744 AJ928744
c	18	59.2	5.3	430	10	CG803196	CG803196 1118039H0
	19	59	5.3	1101	10	CNS000B8	AL063632 Drosophil
c	20	58.4	5.3	1092	10	CNS020K7	AL175696 Tetraodon
c	21	58.2	5.3	1114	1	AJ926009	AJ926009 AJ926009
c	22	57.6	5.2	861	9	BH959723	BH959723 odh52a03.
	23	57.4	5.2	1167	11	CNS07360	AL427102 clone BA0
	24	57.2	5.2	826	10	AG579614	AG579614 Mus muscu
c	25	57.2	5.2	829	10	BX173672	BX173672 Danio rer
	26	57	5.1	1348	10	CG749499	CG749499 P043-4-A0
	27	56.4	5.1	1080	1	AJ925795	AJ925795 AJ925795
	28	56.4	5.1	1190	10	CNS02ON7	AL206908 Tetraodon
	29	56.4	5.1	2087	10	AG333887	AG333887 Mus muscu
c	30	56.2	5.1	471	10	CG807194	CG807194 1118078H0
c	31	56.2	5.1	505	10	CW270542	CW270542 104_742_1
c	32	56.2	5.1	876	10	CNS009G1	AL053529 Drosophil
	33	56.2	5.1	1042	1	AJ927366	AJ927366 AJ927366
c	34	56	5.1	996	10	CNS00FUH	AL071063 Drosophil
c	35	56	5.1	1243	1	AJ928814	AJ928814 AJ928814
c	36	55.6	5.0	716	9	BZ037795	BZ037795 oeh83b12.
c	37	55.6	5.0	1094	10	CNS012FZ	AL101513 Drosophil
	38	55.6	5.0	1391	10	CG754863	CG754863 P050-2-G0
c	39	55.4	5.0	1225	10	CNS0161D	AL106171 Drosophil
	40	55.2	5.0	987	10	CNS014PQ	AL104456 Drosophil
	41	55.2	5.0	1037	10	CW938410	CW938410 TcB18.1_H
	42	55.2	5.0	1225	10	CNS0161D	AL106171 Drosophil
	43	55	5.0	914	9	AZ683059	AZ683059 ENTJU94TF
c	44	54.8	5.0	1022	10	CNS008OJ	AL052236 Drosophil
	45	54.8	5.0	1101	10	CNS0175J	AL107617 Drosophil

ALIGNMENTS

RESULT 1
CNS0039G
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

```
  Query Match             6.9%;  Score 76.6;  DB 10;  Length 1101;
  Best Local Similarity   17.3%;  Pred. No. 7.3e-06;
  Matches  120;  Conservative  284;  Mismatches  291;  Indels    0;  Gaps    0;

Qy          12 GAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAAAGTATATTAAA 71
               | |:|   |:|:: ::| | || ||: | :|||| |||:   : :|||| |:: |||
Db         392 GGGMATATAWAWWWWTTTTTTTTAWAAAWAAAATAATTWWAAWAWAAAAAATTWWAAAA 451

Qy          72 AGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTTAATGAGCAATA 131
               :  ||: :||::      :|       |    |:|  ||  :|  |:|||:|| :   :::
Db         452 WAAAAWAWTAWWTTTAWTWAAAAAAAAAAAATTWTTTTTTTWTTTAWTTWATAWWTTWWW 511

Qy         132 CTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTAA 191
               || | | |       | | |:|| :||| | : :  ::| :|    :    |: :  :
Db         512 TTAAAWAAAAAAAAAAAWAAAWAAAWATAAATWTWWTWWTTYTTWAAWATAAAMCMAAWY 571

Qy         192 GAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCAT 251
               : :| : ::::|::|: :|::::|  |: :|   ::: :  :  :::  ::   :  ::
Db         572 YHTYTTYTYHYYTYWTYTMTWHYHTMYTHAWAHTTWYHWYHTYAMWHWMTWHWTMWAWWH 631

Qy         252 GCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCA 311
               :   :::::       ::  :::|:|:|| :|    ::   :::||:: :  ::  ::
Db         632 WTTYTAAYYYYYTCMYYYHYMHWHHAHAHAAAWWTTHTWWTHAYHWATYHYYYYMYCAMM 691

Qy         312 ACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTT 371
               :  :  ::|: ::::: |::|:|:: :|: : :  ::   :  :  |   :: : ::::
Db         692 CMCTHTCHHCYYYYHHYTAHHTHTHHWYAHYYMWYYMWAYYWMYCTACTYHYHHHHHYHW 751

Qy         372 TAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGA 431
               : |: |: : :::::   : || |: |:  |: ::: | ::::|::  ::::| |:
Db         752 AYHTTWYAWAHAMWMWHHAHYAAAAAWAAWATTHHYHHTTHYMHHTYMYHYYMYTCCYMC 811

Qy         432 TGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATG 491
               | : ::: :   | :|   :: : :|:: :: :::   :: : :|: : ::: :   :
Db         812 TYHCWHYYHTAYTCWTWTHHWMWTWTHWYHHTWWHHTTTHWAWWHTHTWCWWWWHATTWT 871

Qy         492 CCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAA 551
               :|:|   ::::   :: :: :  : :::    : ::: :|  :::: ::: | : :: :
Db         872 WATHCWACMTMHWHHWMHMHHHHMACHAHHHTHMCMCHHHMHCTCHHHHTMYHMTCHWMW 931

Qy         552 GGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAG 611
               :::::::::  : ||   : ::    :   :::: ::: ::| : :  |: : ||:
Db         932 HMHWHWHHWWMATWMTTMTTMMMMCCMMHHHCHMYHMMHMYMYCCHYYCTCHTHATTHYH 991

Qy         612 GGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGCACGGGAGCTAC 671
               |  ::   ::|::||:: |:||:|:|  | ::  ::: ::| :::  :    |
Db         992 YMCTCYHYCTWHTYWTAYWWAWTAHAMTTATWWWWMHWWAHWATWWWWWWWWATAWWACT 1051

Qy         672 TCAAGATTATCGATTCCCTGGATATTTTAATGAAT 706
               ::: ::: :| ::::: :  ::|::  ::: :::
Db        1052 CHHWTWYHTHCTWYYHHTYHMWWAWWMAWHWHHMY 1086


RESULT 2
CW952279/c
LOCUS       CW952279                 891 bp    DNA     linear   GSS 20-DEC-2004
DEFINITION  TcB38.1_B02_T7 Tribolium BAC library Tribolium castaneum genomic,
            genomic survey sequence.
ACCESSION   CW952279
VERSION     CW952279.1  GI:56735316
KEYWORDS    GSS.
SOURCE      Tribolium castaneum (red flour beetle)
  ORGANISM  Tribolium castaneum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
            Tenebrionidae; Tribolium.
REFERENCE   1  (bases 1 to 891)
  AUTHORS   Savard,J. and Tautz,D.
  TITLE     Tribolium castaneum BAC-ends sequencing project
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Savard, J.
            Abteilung fur Evolutionsgenetik, AG Tautz
            Institut fur Genetik, Universitat zu Koln
            Weyertal 121, 50931 Koln, Germany
            Tel: 49 221 470 6911
            Fax: 49 221 470 5975
            Email: savard@uni-koeln.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .891
                     /organism="Tribolium castaneum"
                     /mol_type="genomic DNA"
                     /strain="GA-2"
                     /db_xref="taxon:7070"
                     /clone_lib="Tribolium BAC library"
                     /note="Vector: pBACe3.6; Site_1: EcoR1; Site_2: EcoR1;
                     Library constructed by Exelixis Inc."
ORIGIN

  Query Match             6.4%;  Score 71;  DB 10;  Length 891;
  Best Local Similarity   45.3%;  Pred. No. 0.00011;
  Matches  296;  Conservative    0;  Mismatches  355;  Indels    2;  Gaps    1;

Qy           1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTT--TA 58
               |||  |  |  || || | ||||| ||| |  ||||| |  | ||   | |  | |  ||
Db         826 ATTAATAGATAGATAAGTAATATATATTATAGTATAATATTATAATATTATAATATAATA 767

Qy          59 AAAGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGAT 118
               | |  ||| ||| |  |||  |||        ||      |   |  | |       ||||
Db         766 ATATAATAATAATAATAATAATAATATAATAATATAATAATTATAATAATAATAATAATA 707

Qy         119 TTAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCT 178
               ||||  |   |  |  |||   ||       ||     | |||||||   | |  |||  ||
Db         706 ATAATAATAATAATAATAATAATAATGATAGTAATAATAATAATAATAATAATAATAATA 647

Qy         179 TTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTA 238
               || ||| | ||| |    |      | ||    ||  | |||  ||   ||  | ||  ||
Db         646 GTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 587

Qy         239 TTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATA 298
               |  |  |   |    |  |  |       |||  |||||  ||      |  |  | ||||
Db         586 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 527

Qy         299 TTTATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAA 358
               | || |   | |  |   | |  ||  |     |  || |||      | |   || | |
Db         526 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 467

Qy         359 GGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATT 418
               |||     |  |||     |     |||     ||| |||| || |  ||||     ||
Db         466 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 407

Qy         419 CAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTC 478
               || ||   ||   |||||  ||  |   ||| ||  |   ||      ||  |  |
Db         406 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 347

Qy         479 CATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAAT 538
               || |      | |    ||         | || |   |  |||||  |   ||
Db         346 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 287

Qy         539 ATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATG 598
               |||||| | | ||  || ||  ||  | |   ||| |   |   ||  ||||  || ||
Db         286 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 227

Qy         599 CAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTAT 651
               | |  |  ||       ||| | | ||| | |  |||||  | ||  | ||
Db         226 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 174


RESULT 3
CNS00EVL/c
LOCUS       CNS00EVL                 1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
```

```
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             6.1%;  Score 68;  DB 10;  Length 1101;
  Best Local Similarity   32.8%;  Pred. No. 0.0005;
  Matches  198;  Conservative  104;  Mismatches  299;  Indels    3;  Gaps    1;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              |:|  ||::   | | :| :::|| : | | |||:   : ||||   |::|   | | |
Db       1065 AWTAATTWWTWTATATWTAWWWTATWATATATTAWTTTWTAAAATATAWWTTAATAAATA 1006

Qy         61 AGTATATTAAAAGCAAT---TTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGA 117
              |:||:||::|   ||   | :||: |   |:     |  ||  || : :|| |||| :
Db       1005 TATWTAWTAWWATATATAWATAWAAWTATAWAWTWWAATWAATWWATAWAWTTTAATAAW 946

Qy        118 TTTAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACC 177
              || : :   :|  | |      |   |:|  : | : :  ||||  :    ||  : :
Db        945 AATATWAWWWTWTTATTTTTTTTTTTTTWTTTWTATWTWTTTATATWWTWWAAAAWAWWA 886

Qy        178 TTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTT 237
               :|:|      |:  |    |    || : |  | |:|: ||  :     :  :    :|
Db        885 AWTWTATTTTTTWTTATWATTWATATTAWATTTTTTWTWTATTWWTATATWATWWTATWT 826

Qy        238 ATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTAT 297
              :|: ||  |:   :   |  | :|      | |:|  :|| |:::  :   ::|    :|
Db        825 WTWTTTWATWTATWTATTTATWTWTWTTWTATTWTATWTATAWWWTWWATWWWTAWAAWT 766

Qy        298 ATTTATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAA 357
              || ||| : | |  :  : : ||: |:|   :|||| | | |   | :| |:  :  :
Db        765 ATATATWWATATTATWWTWATTWTTWTTTTATTWTTTTAAWTTATWTATWATTWTTWATW 706

Qy        358 AGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACAT 417
                | | ||||||: ||| :   | | ||     :: |||: : |  |||||  : |
Db        705 TTTTTTTTTTTTTTWTATTTWAAATTATTTAATTAWWWAATWAWTTATTTAATAAWTAA 646

Qy        418 TCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTT 477
              | |  | |    |     ||:|| |  |      ||||  |   :|: |  ||    |
Db        645 TTAAATTATTAAAAATTTTATWAAAAAATATTTTTTTATATTAATWAWATTTATAWATTA 586

Qy        478 CCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAA 537
              || :: |  |     | |||: |  :  ||   | ||:  | |  | |:  :    :|
Db        585 ATATAWWTTWTAAATTWTTAATWTATTWWTTTAAATTATWTTTTTTTTTAWTAWAWWTWA 526

Qy        538 TATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCAT 597
              |:|||: : : :|  | |::::|:|  ||:   :  ||:|||    : |:|   ::  :
Db        525 AAWAATWAWWWWWAAAAATWWWWTWTTTTTWATTWATAAWTTTWTTTWWAWATTWWKKKK 466

Qy        598 GCAA 601
              : |:
Db        465 KKAW 462


RESULT 4
CNS0039G/c
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.9%;  Score 65.4;  DB 10;  Length 1101;
  Best Local Similarity   18.2%;  Pred. No. 0.0018;
  Matches  119;  Conservative  285;  Mismatches  244;  Indels    5;  Gaps    3;

Qy        317 GTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAAT 376
              :: : |::  |  :::::|:: |::: :|:::::||:  ::| ::  || :| |::::::
Db       1097 RWGDDTWDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWWW 1038

Qy        377 TGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAA 436
              : ||  | ::::    |||  :||:| :::| :: : |  ::  ||: : :|  :|:|  :
Db       1037 WWATWDTWWDKWWWWATAAKTDTAWT-WWRTAWRADWAGRDRGAGKRDRDAATDADGAGR 979

Qy        437 ATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTC 496
              ::  :::::::::: ::::  :   ||  | : | : ::  ::::::  |::
Db        978 RDGGRKRKDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKD---RK 922
```

```
Qy        497 AATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATT 556
              |:::: | ::::: : ::|::: : |:::::::::  : |: |: :||: | || ::::
Db        921 ADDDDGAGDKDDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATW-AWAATDWWW 863

Qy        557 ATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCAT 616
              : : :|:::|::  |:: :|:: :: :|: :::|: :  :|  :: :|    ::||
Db        862 WGWADADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKR 803

Qy        617 TTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGCACGGGAGCTACTCAAG 676
              :: :|:::::| :::::|  :|  |  |  :|:::::|:|  :   ::  :: : :::
Db        802 RDRKRADDKRDAADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDR 743

Qy        677 ATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCAAAAAAGCTCGGAATTG 736
              :   | |:::   :::::   :|::::|: ::  :::   :| : |:| : :  |:   :
Db        742 DRAGTAGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRR 683

Qy        737 AATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAAATTGAATTAATCAAAA 796
              ::  ::: : :: : |::   : :|:::   ::: :  :: :: | :||::::|  |::
Db        682 RDRATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWD 623

Qy        797 ATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCTGGAGGGGGGGTAACAT 856
              : :  ::|: ::::::| :   :|   ::: :|:    ::   : |: ::: |    |
Db        622 WAKWDWKTRADRWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTT 563

Qy        857 TTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATAGATGTCAATAAAGAAG 916
              |  :    |    : ::      | : ||:|  : | |    |  :| | |:::|| :
Db        562 TATWTTWAARAAWWAWWAWATTTATWTTTWTTTWTTTTTTTTTTTWTTTAAWWWAAWWTA 503

Qy        917 TTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTATTCATTAAAT 969
              |:||:|      |   |:| || |||        :|: |  ::||:|  || : |
Db        502 TWAAWTAAAWAAAAAAAWAATTTTTTTTTTTTWAWTAAAWWTAWTWTTTTWTT 450
```

RESULT 5
CG807194
LOCUS CG807194 471 bp DNA linear GSS 10-NOV-2003
DEFINITION 1118078H01.y1 1118 - RescueMu Grid S Zea mays genomic, genomic survey sequence.
ACCESSION CG807194
VERSION CG807194.1 GI:38244754
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 471)
AUTHORS Walbot,V.
TITLE Maize genomic sequences found using engineered RescueMu transposon
JOURNAL Unpublished (2001)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 1118078 column: 2
Class: transposon-tagged.

FEATURES Location/Qualifiers
source 1. .471
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="1118 - RescueMu Grid S"
/note="Organ: leaf; Vector: RescueMu (engineered from pBlueScript backbone); Site_1: BamHI; Site_2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescueMu.' Grid S was grown at San Diego in 2002. DNA was extracted from leaf strips, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

ORIGIN

Query Match 5.8%; Score 64.4; DB 10; Length 471;
Best Local Similarity 46.1%; Pred. No. 0.0028;
Matches 215; Conservative 0; Mismatches 251; Indels 0; Gaps 0;

```
Qy        145 TTTACACAAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTAAGAGCTCTGCCTTT 204
              | ||  | |||||||   | |  |||  ||    || ||| | ||| |     |     |
Db          2 TATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 61

Qy        205 TATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTCG 264
              ||    ||  | ||| ||   ||  | ||  || | |  | |   |   | | |
Db         62 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 121

Qy        265 GGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAACACTGTTACTAT 324
                |||   ||||| ||    |  | |  | ||| | || |   | |  | | | | |||
Db        122 AATAATAATAATAATAATAATTATAATAATAATAATAATAATAATTATAATTATAATTAT 181

Qy        325 CCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTC 384
              ||    || || |||   | ||  ||    | |    || ||| |  |
Db        182 AATTATAATTATAATAATAATTATATTTATAGTTATATTTATAATTATAATTATAATTAT 241

Qy        385 AATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATAT 444
              ||||  | | |||| || |  ||||   |||  ||  ||   ||    |||| |||
Db        242 AATTATAATTATAATAATAATTATAATTATAATTATATTTATAATTATAATTATAATTAT 301

Qy        445 GTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAA 504
              ||   |||||  |   ||     ||  |  |     || |   |  | |    ||
Db        302 ATTTATAATTATAATTATAATTATAATAATAATTATAATAATAATTATAATTATAATAAT 361

Qy        505 TGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATA 564
                | || |   |  ||||| |   ||  |||||| | | ||  || ||  ||
Db        362 AATAATAATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 421

Qy        565 ATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTA 610
              | |   ||| |   |   || |||||  || ||  | |  |  ||
Db        422 AATAATAATAATAATAATAATCATAATAATAATAATAATAATAATA 467
```

RESULT 6
BX139987
LOCUS BX139987 844 bp DNA linear GSS 13-MAR-2003
DEFINITION Danio rerio genomic clone DKEY-99E7, genomic survey sequence.
ACCESSION BX139987
VERSION BX139987.1 GI:27971314
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 844)
AUTHORS Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Unpublished
COMMENT This sequence was generated from the T7 end of BAC 99E7. 99E7 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_rerio/.

FEATURES Location/Qualifiers
source 1. .844
/organism="Danio rerio"

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-99E7"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.7%;  Score 63.6;  DB 10;  Length 844;
  Best Local Similarity   45.4%;  Pred. No. 0.0043;
  Matches  228;  Conservative    0;  Mismatches  274;  Indels    0;  Gaps    0;

Qy        144 CTTTACACAAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTAAGAGCTCTGCCTT 203
              | |||    | |||||||  || |  |||  |      || ||| | | |       |
Db        343 CATTAATAATAATAATAATGATAATAATAATGATAATAATAATAATGATGATAATAATAA 402

Qy        204 TTATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTC 263
              | ||    ||  | ||  ||   ||  | ||  |  |  |  |    || |  |
Db        403 TAATATTGATAATAATGATAATAATAATAATAATGATAATAATAATAATATTGATAATAA 462

Qy        264 GGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAACACTGTTACTA 323
              |  |||   ||||| || |      | |  | ||||| || |  || |  |   | | |
Db        463 TGATAATAATAATAATAATGATAATAATAATAATATTGATAATATTGATGATAATAATAA 522

Qy        324 TCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCT 383
              |  |      |  || |||    || |   || |    |||      |  |||    |
Db        523 TAATGATAATAATAATAATAATGATAATAATAATGATAATAATAATAATAATGATAATAA 582

Qy        384 CAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATA 443
               |||   | | | |||| || || | ||   ||     || || | ||   |||||  |
Db        583 TAATAATGATGATAATAATAATGATGATAATAATGATAATAATAACAATAATAATAATAA 642

Qy        444 TGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAA 503
              || |   ||| ||  |     |     || |  |     || |        | |    ||
Db        643 TGATAATAATAATAATAATGATGATAATAATAATAATAATAATGATGATAATAATAATAA 702

Qy        504 ATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTAT 563
                | || ||  |  ||||| || |||  |||||| | | || ||| ||  ||
Db        703 TGATACTAATAATGACAATAATAATAATGACAATAATAATAATAATAATGATAATAATAA 762

Qy        564 AATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAAT 623
                |  ||| |   |   ||  |||| ||| ||   | || |  ||        | |
Db        763 TGATAATAATAATGATACTAATAATAATGATAATAATAATGATAATAATAATAATGATGA 822

Qy        624 TTATACTGAATATAATCCTGAT 645
              | ||| | |  || ||  | ||
Db        823 TAATAATAATAATGATGATAAT 844


RESULT 7
CW973509
LOCUS       CW973509                889 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  AIAA-aaa59f06.b1 Ancylostoma caninum whole genome shotgun library
            (AIAAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION   CW973509
VERSION     CW973509.1  GI:56776238
KEYWORDS    GSS.
SOURCE      Ancylostoma caninum (dog hookworm)
  ORGANISM  Ancylostoma caninum
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
            Ancylostomatoidea; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   Mitreva,M., McCarter,J.P., Pape,D., Ritter,E., Tsagareishvili,R.,
            Ronko,I., Martin,J., Wylie,T., Dante,M., Meyer,R., Messina,D.,
            Waterston,R.H., Clifton,S.W. and Wilson,R.
  TITLE     Genome Survey sequences from the parasitic nematode Ancylostoma
            caninum
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Mitreva M
            Washington University in St. Louis
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: nematode@watson.wustl.edu
            Genomic DNA provided by John Hawdon (mtmjmh@gwumc.edu) DNA
            sequenced by Washington University Genome Sequencing Center
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .889
                     /organism="Ancylostoma caninum"
                     /mol_type="genomic DNA"
                     /strain="Baltimore"
                     /db_xref="taxon:29170"
                     /dev_stage="Adult"
                     /lab_host="GS10"
                     /clone_lib="Ancylostoma caninum whole genome shotgun
                     library (AIAAGSS 001)"
                     /note="Vector: pOTW13; Site_1: BstX1; Site_2: BstX1;
                     Ancylostoma caninum genomic DNA was randomly sheared,
                     end-repaired and size fractioned to enrich for 2-4 kb
                     fragments. Genomic DNA was provided by John Hawdon
                     (mtmjmh@gwumc.edu) at George Washington University.
                     Sequencing by Washington University Genome Sequencing
                     Center, St. Louis, MO."
ORIGIN

  Query Match             5.7%;  Score 62.6;  DB 10;  Length 889;
  Best Local Similarity   43.6%;  Pred. No. 0.0071;
  Matches  325;  Conservative    0;  Mismatches  419;  Indels    1;  Gaps    1;

Qy        112 AATAGATTTAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAAT 171
              ||||   |||| |  |  |  |||    |  |  ||   | |||||||      |  |||
Db         23 AATAATAATAATAATAATAATGATAATCCTAACAATAATAATAATAATAGTAATAAT 82

Qy        172 CTTACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGT 231
              ||    || ||| | ||| | |  |    | ||    ||  | ||| ||   ||  |
Db         83 AATAATAATAATAATAATAATAACAGTAATAATAATAATAATAATAATAATAATAATAAT 142

Qy        232 GATGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATT 291
              ||  || |  |  |    |  ||  |  |      |||   ||||| ||       | |
Db        143 AATAATAATAATAATAATAACAGTAATAATAATAATAATAATAATAATAATAATAATAAT 202

Qy        292 GGTTATATTTATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATA 351
              | ||| | ||  |  |  |  |  |  ||  |      |  || |||    | |
Db        203 AATAATAATAATAATAATAATAATAATAATGATAATAATAATAATAATAATAATAATAAT 262

Qy        352 GATCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAAT 411
              || | |  |||     |  |||    |   ||||     ||| |||| || |  ||||
Db        263 AATAATAATAATAATAATAATAATAATAATAACAATAATAATAATAATAATAATAATAAT 322

Qy        412 GCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACAT 471
                ||    || ||   ||  |||||  ||  |   ||| ||  |    |      ||
Db        323 AATAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATGATAATAAT 382

Qy        472 TCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATT 531
              |      || |      | |   ||          | ||  |   |  |||||  |
Db        383 AATAACAATAATAATAATAATAACAATAATAATAATAATAATAATAATAATAATAATAAT 442

Qy        532 GCCAAATATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAA 591
                ||  |||||| | |  |  || ||  ||  | |  |||  |   |   ||  ||||
Db        443 AATAATAATAATAATAATGATAATAATAATAACAATAATAATGATAATAATAATAATAAT 502

Qy        592 GATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTAT 651
              || ||   | |  |  ||        ||| | | ||| | |  |||||  |  |  | ||
Db        503 AATAATAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAGTAATAAT 562

Qy        652 TTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGA-ATTGAT 710
                    |       || | |  ||          |    ||  |||| | | | ||
Db        563 AATAATAATAATAATAATAATAACAATAATAATAATAATAATAATAATAATAATAAT 622

Qy        711 GGTTTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAA 770
              |  | | ||  |   |||   |  | ||| |  ||| ||       |||   |
Db        623 AATAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 682

Qy        771 ACTTGAACAAATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGA 830
              | | | |  |||   |  ||| | ||  |   || || | ||     | | | |
```

```
Db        683 AATAGTAGTAATAATAACAATAATAATATATTTTACTATTATTATTACTAGCGTTCGCTG 742

Qy        831 AGGCGGGCCTGGAGGGGGGGTAACA 855
              ||   ||  ||| | ||| | | |
Db        743 CGGGACGCGAGGAAGTGGGATGAAA 767


RESULT 8
AQ946120/c
LOCUS       AQ946120                 641 bp    DNA     linear   GSS 27-JAN-2000
DEFINITION  Sheared DNA-46J23.TR Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-46J23, genomic survey sequence.
ACCESSION   AQ946120
VERSION     AQ946120.1  GI:6769385
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
            Donelson,J., Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: Sheared DNA-46J23.TF
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Reverse
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Trypanosoma brucei"
                     /mol_type="genomic DNA"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-46J23"
                     /clone_lib="Sheared DNA"
                     /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from a cloned population of
                     Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                     sheared to give a tight size distribution (approx 2 kb).
                     The v + i method used for the library construction is
                     described in detail in Smith, H.O. and Venter, J.C.
                     (Making small insert libraries for whole genome shotgun
                     sequencing projects.  In Genome Sequencing: A Practical
                     Approach, eds. M. Vaudin and B. Barell, Oxford University
                     Press, 1999)."
ORIGIN

  Query Match             5.6%;  Score 62;  DB 9;  Length 641;
  Best Local Similarity   45.0%;  Pred. No. 0.0093;
  Matches  233;  Conservative    0;  Mismatches  285;  Indels    0;  Gaps    0;

Qy        134 ATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTAAGA 193
              ||||   | |   ||   | |||||||   | |  |||  ||    || ||| | ||| |
Db        619 ATAGTACTACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 560

Qy        194 GCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGC 253
                  |     | ||    ||  | ||| ||   ||  | ||  || |  |  |  |
Db        559 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 500

Qy        254 GTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAAC 313
              | |  |       |||   ||||| ||       | |  | ||| | || |   | |
Db        499 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 440

Qy        314 ACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTA 373
              | | |  ||  |     |  || |||     | |   || | |  |||     |  ||
Db        439 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 380

Qy        374 AATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATG 433
              |    |   |||     ||| |||| || |  ||||    ||     || ||   ||
Db        379 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 320

Qy        434 CAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCC 493
              |||||  ||  |  ||| ||  |  ||    ||  |  |    ||
Db        319 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 260

Qy        494 CTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGG 553
              | |   ||          | || |   |  |||||  |    ||  ||||||  | | ||
Db        259 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 200

Qy        554 ATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGG 613
              || ||  ||  | |   ||| |   |   ||  |||  || ||   | |  |  ||
Db        199 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 140

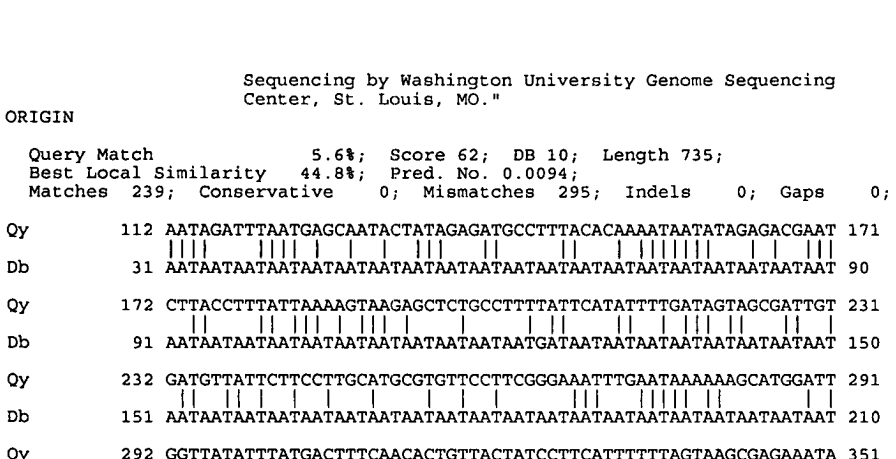
Qy        614 CATTTAAAATTTATACTGAATATAATCCTGATGTTTAT 651
              ||| | | ||| | |  |||||  ||||| | ||
Db        139 ATAATAATAATAATAATAATAATAATGATGATGAT 102


RESULT 9
CW960936
LOCUS       CW960936                 735 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  AIAA-aac14d08.b1 Ancylostoma caninum whole genome shotgun library
            (AIAAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION   CW960936
VERSION     CW960936.1  GI:56761663
KEYWORDS    GSS.
SOURCE      Ancylostoma caninum (dog hookworm)
  ORGANISM  Ancylostoma caninum
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
            Ancylostomatoidea; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Mitreva,M., McCarter,J.P., Pape,D., Ritter,E., Tsagareishvili,R.,
            Ronko,I., Martin,J., Wylie,T., Dante,M., Meyer,R., Messina,D.,
            Waterston,R.H., Clifton,S.W. and Wilson,R.
  TITLE     Genome Survey sequences from the parasitic nematode Ancylostoma
            caninum
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Mitreva M
            Washington University in St. Louis
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: nematode@watson.wustl.edu
            Genomic DNA provided by John Hawdon (mtmjmh@gwumc.edu) DNA
            sequenced by Washington University Genome Sequencing Center
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Ancylostoma caninum"
                     /mol_type="genomic DNA"
                     /strain="Baltimore"
                     /db_xref="taxon:29170"
                     /dev_stage="Adult"
                     /lab_host="GS10"
                     /clone_lib="Ancylostoma caninum whole genome shotgun
                     library (AIAAGSS 001)"
                     /note="Vector: pOTW13; Site_1: BstX1; Site_2: BstX1;
                     Ancylostoma caninum genomic DNA was randomly sheared,
                     end-repaired and size fractioned to enrich for 2-4 kb
                     fragments. Genomic DNA was provided by John Hawdon
                     (mtmjmh@gwumc.edu) at George Washington University.
```

Sequencing by Washington University Genome Sequencing Center, St. Louis, MO."

ORIGIN

Query Match 5.6%; Score 62; DB 10; Length 735;
Best Local Similarity 44.8%; Pred. No. 0.0094;
Matches 239; Conservative 0; Mismatches 295; Indels 0; Gaps 0;

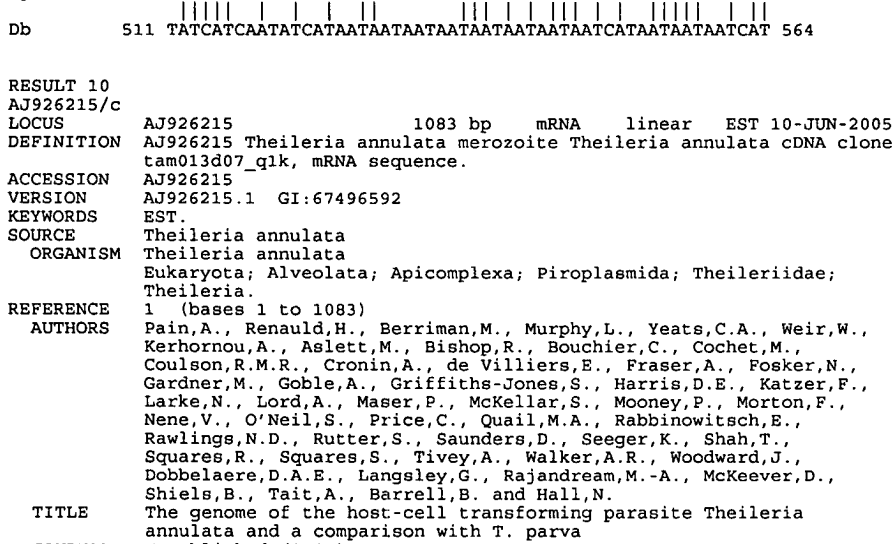


RESULT 10
AJ926215/c
LOCUS AJ926215 1083 bp mRNA linear EST 10-JUN-2005
DEFINITION AJ926215 Theileria annulata merozoite Theileria annulata cDNA clone tam013d07_q1k, mRNA sequence.
ACCESSION AJ926215
VERSION AJ926215.1 GI:67496592
KEYWORDS EST.
SOURCE Theileria annulata
ORGANISM Theileria annulata
Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae; Theileria.
REFERENCE 1 (bases 1 to 1083)
AUTHORS Pain,A., Renauld,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W., Kerhornou,A., Aslett,M., Bishop,R., Bouchier,C., Cochet,M., Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Fosker,N., Gardner,M., Goble,A., Griffiths-Jones,S., Harris,D.E., Katzer,F., Larke,N., Lord,A., Maser,P., McKellar,S., Mooney,P., Morton,F., Nene,V., O'Neil,S., Price,C., Quail,M.A., Rabbinowitsch,E., Rawlings,N.D., Rutter,S., Saunders,D., Seeger,K., Shah,T., Squares,R., Squares,S., Tivey,A., Walker,A.R., Woodward,J., Dobbelaere,D.A.E., Langsley,G., Rajandream,M.-A., McKeever,D., Shiels,B., Tait,A., Barrell,B. and Hall,N.
TITLE The genome of the host-cell transforming parasite Theileria annulata and a comparison with T. parva
JOURNAL Unpublished (2005)
COMMENT Contact: Pain A
The Pathogen Sequencing Unit
The Wellcome Trust Sanger Institute
Genome Campus, CB10 1SA, UNITED KINGDOM
Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of Veterinary Infection and Immunity, ICM, University of Glasgow, UK.

FEATURES Location/Qualifiers
source 1. .1083
/organism="Theileria annulata"
/mol_type="mRNA"
/isolate="Ankara (clone D7)"
/db_xref="taxon:5874"
/clone="tam013d07_q1k"
/dev_stage="merozoite"
/lab_host="Bos taurus (cow)"
/clone_lib="Theileria annulata merozoite"
/note="country: Turkey:Ankara"

ORIGIN

Query Match 5.5%; Score 60.8; DB 1; Length 1083;
Best Local Similarity 38.5%; Pred. No. 0.017;
Matches 259; Conservative 0; Mismatches 412; Indels 1; Gaps 1;



RESULT 11
CNS00EVL
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999

```
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             5.5%;  Score 60.6;  DB 10;  Length 1101;
  Best Local Similarity   33.2%;  Pred. No. 0.019;
  Matches  213;  Conservative  107;  Mismatches  312;  Indels    9;  Gaps    2;

Qy          4 CTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAAAGT 63
              :|:| ::: : :|||  :: | :   : |  :|| : ||||:  :::   |||::::: :
Db        460 HTWTMMMMMMWWAATWTWWAAAWAAAWTTATWAATWAAAAAAWAWWWWATTTTTWWWWWTW 519

Qy         64 ATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTTAAT 123
              || :|  :|   :| :|||| |   |:|     |  |    | |  |||||: |  | :||:
Db        520 ATTWTTTWAWWTWTAWTAAAAAAAAAWATAATTTAAAWWAATAWATTAAWAATTTAWAAW 579

Qy        124 GAGCAATACTATAGAGATGCCTTTACACAAAATAA--------TATAGAGACGAATCTTA 175
                | || |:|| | ||    |||    ||| ||        |:   | |    |  ||
Db        580 WTATATTAATWTATAAATWTWATTAATATAAAAAAATATTTTTTWATAAAATTTTTAATA 639

Qy        176 CCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATG 235
              || ||||:   | | |    |   |:::     | |    |:  |   |      |
Db        640 ATTTAATTAWTTATTAAATAAWTWATTWWWTAATTAAATAATTTWAAATAWAAAAAAAAA 699

Qy        236 TTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTT 295
                |   :       ||       |  :|    |:| |  ||:||:||   ::  | |  ::
Db        700 AAAAAAWATWAAWAATWATAWATAAWTTAAAAWAATAAAAWAAWAATWAWWATAATATWW 759

Qy        296 ATATTTATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATC 355
              |||| ||:   | : :| :   ::  ||:   : || :  :|  :|   | |:||| ||
Db        760 ATATATAWTTWTAWWWATWWAWWWTATAWATAWAATAWAAWAWAWATAAATAWATAWATW 819

Qy        356 AAAGGAATGTGTTTTT-TAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCA 414
              |||  :|: | :: |: || |:: ||    ||       ||| | :| |: |  :||   |
Db        820 AAAWAWAWATAWWATWATATAWWAATAWAWAAAAAATWTAATATWAATWATAAWAAAAAA 879

Qy        415 CATTCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCT 474
              |:    :: :|    :   | |||||:|: |    |:  |      ||||  | |:: |
Db        880 TAWAWTTWWTWTTTTWWAWWATATAAAWAWATAWAAAWAAAAAAAAAAAAAATAAWAWWWT 939

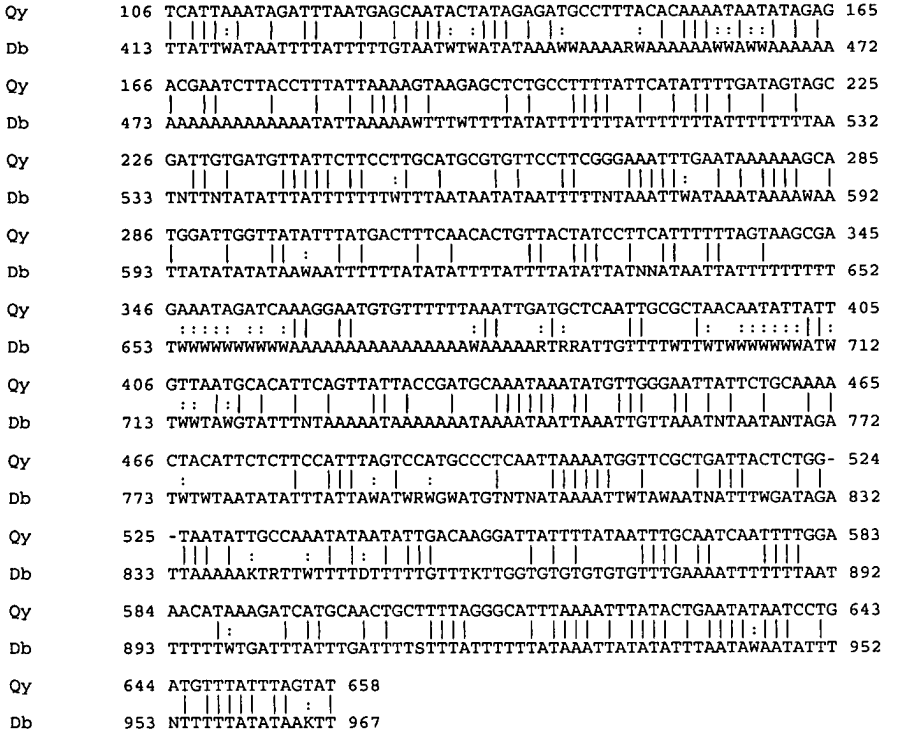
Qy        475 CTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCC 534
                |  :||| |  |:    | :|||:|  :  :|   | :||:   :  ||   ::
Db        940 WATATTWTTATTAAAWTWTATWWATTWATTWWAWTWTATAWTTWTATWTATATATWWTAW 999

Qy        535 AAATATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGAT 594
              |:    || |   |:  ||  ||| : ||:|   |   :|| ::  |:| |||:| :
Db       1000 TAWATATATTTATTAAWWTATATTTTAWAAAWTAATATATWATAWWWTAWATATAWAWWA 1059

Qy        595 CATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATA 635
              |      |   |:||    |:: ||||:  |:|   |::|
Db       1060 ATTAWTTATATATWTAAWATAWWAAAAAWAAAWAWATAWWA 1100


RESULT 12
CNS02T50
LOCUS       CNS02T50                1027 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            163M16 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL212733
VERSION     AL212733.1  GI:7871552
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1027)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1027
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="163M16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG163BG08LP1
                     end : T7"
ORIGIN

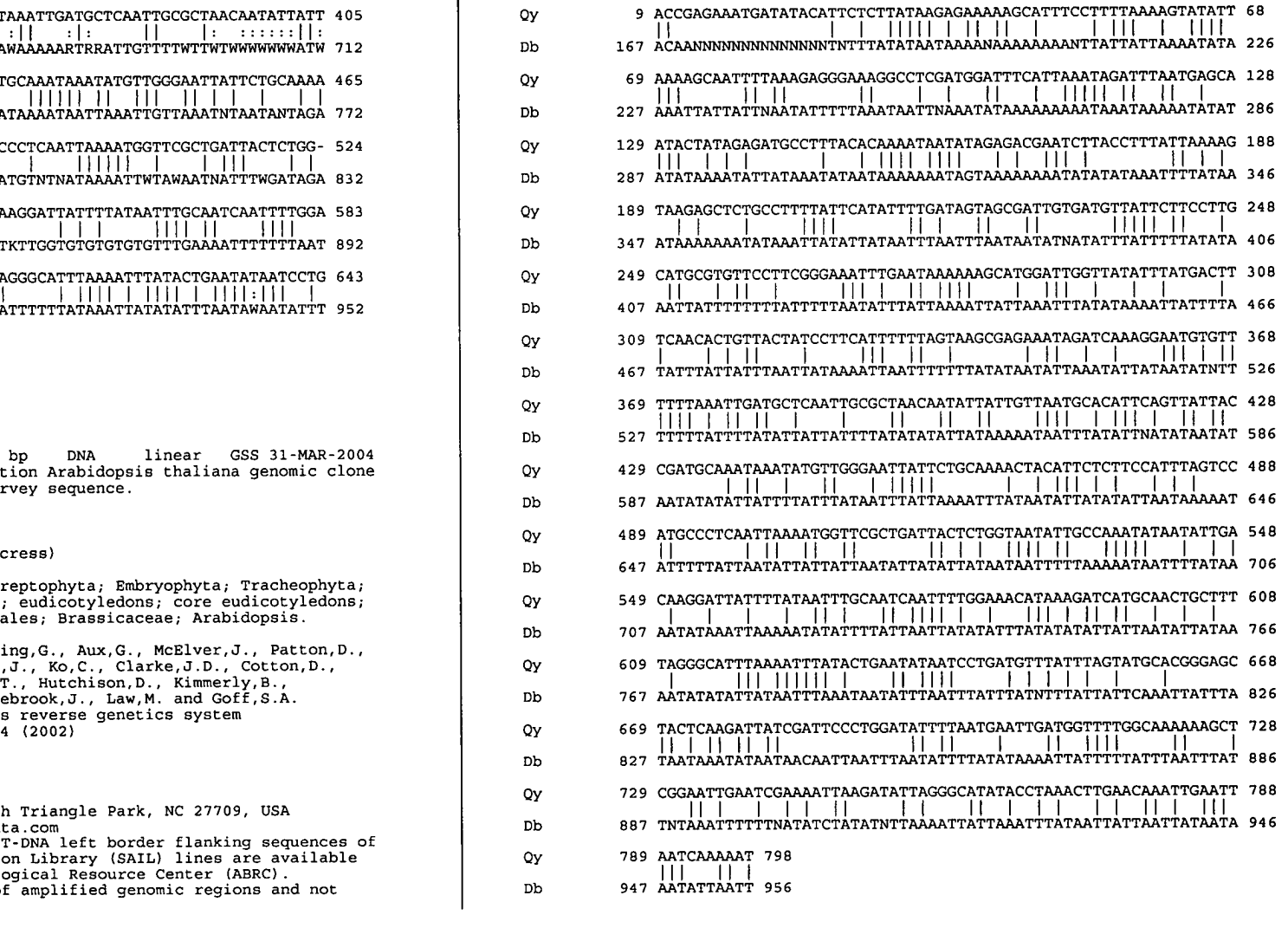
  Query Match             5.4%;  Score 60.2;  DB 10;  Length 1027;
  Best Local Similarity   38.9%;  Pred. No. 0.023;
  Matches  216;  Conservative   43;  Mismatches  294;  Indels    2;  Gaps    1;
```



RESULT 13
CL469446
LOCUS CL469446 969 bp DNA linear GSS 31-MAR-2004
DEFINITION SAIL_1303_G10.v1 SAIL Collection Arabidopsis thaliana genomic clone SAIL_1303_G10.v1, genomic survey sequence.
ACCESSION CL469446
VERSION CL469446.1 GI:45872351
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 969)
AUTHORS Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D., Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D., Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B., Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
TITLE A high-throughput Arabidopsis reverse genetics system
JOURNAL Plant Cell 14 (12), 2985-2994 (2002)
PUBMED 12468722
COMMENT Contact: Sessions A
Applied Trait Genetics
Syngenta Biotechnology Inc.
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
Email: allen.sessions@syngenta.com
ABRC Stock Number CS848789; T-DNA left border flanking sequences of Syngenta Arabidopsis Insertion Library (SAIL) lines are available through the Arabidopsis Biological Resource Center (ABRC).
Sequences represent a pool of amplified genomic regions and not single contiguous sequences.

Class: TDNA tagged.
FEATURES Location/Qualifiers
source 1. .969
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="SAIL_1303_G10.v1"
/clone_lib="SAIL Collection"
/note="T-DNA left border sequences were isolated using a modified TAIL-PCR strategy"
ORIGIN

Query Match 5.4%; Score 60; DB 10; Length 969;
Best Local Similarity 41.0%; Pred. No. 0.026;
Matches 324; Conservative 0; Mismatches 466; Indels 0; Gaps 0;



```
RESULT 14
CC253231/c
LOCUS       CC253231               1277 bp    DNA     linear   GSS 13-MAY-2003
DEFINITION  CH261-180N11_RM1.1 CH261 Gallus gallus genomic clone CH261-180N11,
            genomic survey sequence.
ACCESSION   CC253231
VERSION     CC253231.1  GI:30589981
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1277)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: RM1 TACGACTCACTATAGGGAGA
            Class: BAC ends
            High quality sequence start: 24
            High quality sequence stop: 85.
FEATURES             Location/Qualifiers
     source          1. .1277
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-180N11"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN

  Query Match             5.4%;  Score 60;  DB 9;  Length 1277;
  Best Local Similarity   45.6%;  Pred. No. 0.026;
  Matches  333;  Conservative    0;  Mismatches  385;  Indels   12;  Gaps    3;

Qy        104 TTTCATTAAATAGATTTAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAG 163
              | | | || |||     | || |     || ||||     |      || | || || ||| |
Db       1092 TATTAATATATAATAATTATAATTTTTAATATATTATTATTAATATAGAATATTATAAAA 1033

Qy        164 AGACGAATCTTACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTA 223
                |  | | |||  ||| ||      ||| |  | |   || |||| || ||||    ||
Db       1032 TTATAATTTTTATATTTTTTTTTTATAATAAATTTATTTTATATTTATTTTTTTTATTTA 973

Qy        224 GCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAG 283
                | | || |  ||  | |  |   | ||  |       |||||  | ||| | |
Db        972 TATAAATTAATATATTTTATTATATAAAATATTTTATTATTTTTAATTTAATTAATATAA 913

Qy        284 CATGGATTGGTTATATTTATG---ACTTTCAACACTGTTACTATCCTTCATTTTTTAGTA 340
                |  |||   ||||||||||    | |||  |       |  | |   |  ||||||  ||
Db        912 TTTAAATTATATATATTTATGTTAATTTTATAATTATATTATTTTTATATTTTTTTTATA 853

Qy        341 AGCGAGAAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATA 400
                |    | ||     ||  ||     | | | ||||||  || ||       |    | | | ||
Db        852 ATATAAAATATTATATAATTTTTATATATTTTAATTTTATTTAATTATAATATTATATTA 793

Qy        401 TTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTG 460
                | | || |     |  | ||||     |   ||| |  |  ||   ||  ||  |
Db        792 ATTATATTTAATTTATTAAAATTATATATATTATTAATTATAAAATTTTTAAAAATATTT 733

Qy        461 CAAAACTACATTCTCTTCCATT---TAGTCCATGCCCTCAATTAAAATGGTTCGCTGATT 517
                ||   | ||| | |   |||    | |   |      |||| |  |  ||   |  ||
```

```
Db        732 ATAATTAATATTATATAAAATTAAAAATTAATTTTATATAATTTATTTAATTTATTTTTT 673

Qy        518 ACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCAATCAAT 577
              | | |  |  || |    ||| | ||  | | |   ||| | |  || |||   |   ||
Db        672 ATTATTATTTTAATATTTAATTTTATTATTATATATATTTTATATTATTTTTATAATTAT 613

Qy        578 TTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATA 637
               ||  |||  |||   | ||  || | |||       || || ||| ||| |  | | ||
Db        612 ATTTTAAATTTAATTTTTATTTAAATAATTT------ATATATAATATATTTTTATTTTA 559

Qy        638 ATCCTGATGTTTATTTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATATT 697
              ||    || || | ||| ||      || |   |   |   |  ||   |   || |
Db        558 ATATATATTTTAAATTATTAAAATATATAGTTTAATATTTTATTTATTTAAATAATTAAT 499

Qy        698 TTAATGAATTGATGGTTTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAG 757
              ||||| | ||     |||   | ||||       |   | |    ||||    |||||
Db        498 TTAATAATTTTTATAATTTATAATAAAAATATAATAAATATTATTTAATTTTTTTATTAA 439

Qy        758 GGCATATACCTAAACTTGAACAAATTGAATTAATCAAAAATTGCATTGCTGTAATACAAC 817
                | ||   ||   |  | | || | |     ||  ||    ||    |
Db        438 ATAAAATTATTATTATATTATATTTTTATTATTATAATTATATATTTTTATTATATAATA 379

Qy        818 CAACCTTATT 827
                ||   ||||
Db        378 AAATTATATT 369


RESULT 15
CNS00EO7/c
LOCUS       CNS00EO7               1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069440
VERSION     AL069440.1  GI:4949583
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29P01"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.4%;  Score 59.8;  DB 10;  Length 1101;
```

```
  Best Local Similarity   38.6%;  Pred. No. 0.028;
  Matches  217;  Conservative   69;  Mismatches  268;  Indels    8;  Gaps    3;

Qy          26 ATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAAAGTATATTAAAAGCAATTTTAAAGA 85
               ||| | || | : | |   :  |||| :  |  |:: :|| |:: |  ||   |::
Db        1005 ATTTTTTTTTTMAATATYCWATATTTTMAATACAWWAYATTTWWTATACATAATWWTTTT 946

Qy          86 GGGAAAGGCCTCGATGGATTTCATTAAATAGATTTAATGAGCAATACTATAGAGATGCCT 145
               |    |  :   ||   | : || : | || |: |   ||  | |  | ::   |
Db         945 TTATATACAATTTWAAAATAAAAAWTAACWWAATTTAWAAAWCATTWTTTTTAAWWTAAT 886

Qy         146 TTACACAAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTAA--GAGCTCTGCCTT 203
               ||:   | |:|: ||:: | |  ||| |||  :| | :||||  :|   :  | | : |:
Db         885 TTWATTACAWTWTTAWWTAAAAAAATWTTAAAWTAAAWAAAAARWATTAWAATTTAYATW 826

Qy         204 TTATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTC 263
                |||: | |::|  |:| ||   ::  |  | | :   ||  ||       : :|   |
Db         825 ATATWTAAAWWTATAWATTATTMWWAATWTTATAWNATTTTTTTTWTAWAAWTWTTWATW 766

Qy         264 GGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAACACTGTTACTA 323
               || ||| |:|::  ||  |   |    |||| |||||   ||   : | : ||  |
Db         765 AWTAATTTTAAWTWWTTAAATAAAAAAAATTTATTTTTATTTWTTATTWAAAWTTTTWTT 706

Qy         324 TCCTTCATTTTTTAGTAAGCGA-----GAAATAGATCAAAGGAATGTGTTTTTTAAATTG 378
               |       :::|||| ||:   |       |||:| :| |:   |||   |:| || ||
Db         705 TTWAATTWWYTTTAATAWTTAAAWTWTTAAAWAWWTAAWTTAAATAAATWTATTTAAAAT 646

Qy         379 ATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAAT 438
               :|      ||      ||: ::|  ||:: ||  |  :|:|| | |   ||    :  ||||
Db         645 WTWAAWTATAAATTTAWADWTTATAWWTTTTTTTWWCWTTWAATATATAAAATWAAAAAT 586

Qy         439 AAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCAT-GCCCTCA 497
               ||  | ||   |  |::: |  |||: |: :|| | |   |:: | |   |     | |
Db         585 TAAATTWTTTTTATATWWWATAAAAAMATWWWTTATTTATAAWWAATTATTTAWAWTTGA 526

Qy         498 ATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTA 557
               ||  | |  ||   |  |||   |  || ||| |          ||||:|| :  |   ||
Db         525 TTTTTATTTATTTWTTTTTTAAAWTTTTATTATATWATTTATTAATWTTAWTTATYTTTT 466

Qy         558 TTTTATAATTTGCAATCAATTT 579
               |||||||  || |:|   ||||
Db         465 TTTTATATCTTTCMAAWTATTT 444


Search completed: May 11, 2006, 05:00:33
Job time : 3548.5 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 20:34:42 ; Search time 154.411 Seconds
(without alignments)
12743.667 Million cell updates/sec

Title: US-09-423-093B-2_COPY_5365_6471
Perfect score: 1107
Sequence: 1 attcttttaccgagaaatga.....aaattgaatcccgatcttaa 1107

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	1107	100.0	14187	3	US-09-453-702B-121	Sequence 121, App
c	2	1107	100.0	14187	3	US-10-114-170-121	Sequence 121, App
	3	61.4	5.5	1141	3	US-09-806-708B-22	Sequence 22, Appl
	4	54	4.9	832	3	US-09-621-976-2813	Sequence 2813, Ap
c	5	52.6	4.8	1141	3	US-09-806-708B-22	Sequence 22, Appl
	6	52	4.7	50000	3	US-09-662-254B-23	Sequence 23, Appl
	7	50.8	4.6	640681	3	US-09-790-988-1	Sequence 1, Appli
	8	49.4	4.5	1410	3	US-09-662-254B-33	Sequence 33, Appl
	9	49.4	4.5	50000	3	US-09-662-254B-24	Sequence 24, Appl
	10	47.2	4.3	96922	3	US-09-949-016-17061	Sequence 17061, A
	11	47	4.2	837	3	US-08-998-416-288	Sequence 288, App
c	12	47	4.2	32392	3	US-09-662-254B-27	Sequence 27, Appl
	13	46.8	4.2	3399	3	US-09-601-198-43	Sequence 43, Appl
c	14	45.6	4.1	50000	3	US-09-662-254B-23	Sequence 23, Appl
c	15	45.2	4.1	50000	3	US-09-662-254B-24	Sequence 24, Appl
	16	45	4.1	50000	3	US-09-662-254B-26	Sequence 26, Appl
c	17	44.8	4.0	1881	3	US-09-270-767-5997	Sequence 5997, Ap
c	18	44.8	4.0	1881	3	US-09-270-767-21279	Sequence 21279, A
	19	44.8	4.0	263693	3	US-09-949-016-12386	Sequence 12386, A
	20	44.8	4.0	263694	3	US-09-949-016-16915	Sequence 16915, A
c	21	43.8	4.0	832	3	US-09-621-976-2813	Sequence 2813, Ap
c	22	43.8	4.0	14066	3	US-09-601-198-56	Sequence 56, Appl
	23	43.8	4.0	125902	3	US-09-949-016-13715	Sequence 13715, A
	24	43.6	3.9	18773	3	US-09-949-016-14164	Sequence 14164, A
c	25	43.2	3.9	7218	2	US-08-232-463-14	Sequence 14, Appl
c	26	43	3.9	3057	3	US-09-601-198-55	Sequence 55, Appl
c	27	43	3.9	3069	3	US-09-710-279-4415	Sequence 4415, Ap
c	28	43	3.9	3335	3	US-09-710-279-3366	Sequence 3366, Ap
	29	43	3.9	3456	3	US-09-710-279-4026	Sequence 4026, Ap
c	30	43	3.9	3638	3	US-09-710-279-3640	Sequence 3640, Ap
	31	43	3.9	3737	3	US-09-710-279-4196	Sequence 4196, Ap
	32	43	3.9	4136	3	US-09-710-279-4036	Sequence 4036, Ap
c	33	43	3.9	4346	3	US-09-710-279-3702	Sequence 3702, Ap
c	34	43	3.9	50000	3	US-09-662-254B-25	Sequence 25, Appl
	35	43	3.9	134987	3	US-09-949-016-15348	Sequence 15348, A
	36	43	3.9	134987	3	US-09-949-016-15349	Sequence 15349, A
	37	43	3.9	134987	3	US-09-949-016-15350	Sequence 15350, A
	38	43	3.9	134987	3	US-09-949-016-15507	Sequence 15507, A
	39	43	3.9	134987	3	US-09-949-016-15508	Sequence 15508, A
	40	43	3.9	134987	3	US-09-949-016-15509	Sequence 15509, A
	41	42.8	3.9	615	3	US-08-998-416-186	Sequence 186, App
	42	42.8	3.9	1055	3	US-09-806-708B-23	Sequence 23, Appl
c	43	42.6	3.8	665	2	US-08-883-795A-36	Sequence 36, Appl
c	44	42.6	3.8	187169	3	US-09-949-016-12776	Sequence 12776, A
c	45	42.6	3.8	191569	3	US-09-949-016-15940	Sequence 15940, A

ALIGNMENTS

RESULT 1
US-09-453-702B-121/c
; Sequence 121, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14187
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-09-453-702B-121

```
  Query Match           100.0%;  Score 1107;  DB 3;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 8.8e-239;
  Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8636 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 8577

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8576 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 8517

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8516 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 8457

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8456 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 8397

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8396 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 8337

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8336 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 8277

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8276 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 8217

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8216 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 8157

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8156 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 8097

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8096 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 8037

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8036 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 7977

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7976 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 7917

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7916 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 7857

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7856 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 7797

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7796 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 7737

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7736 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 7677

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7676 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 7617
```

```
Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7616 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 7557

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
              |||||||||||||||||||||||||||
Db       7556 AAACAAGAAATTGAATCCCGATCTTAA 7530
```

RESULT 2
US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Patent No. 6855814
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6855814el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14187
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121

```
  Query Match           100.0%;  Score 1107;  DB 3;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 8.8e-239;
  Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8636 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 8577

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8576 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 8517

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       8516 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 8457

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8456 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 8397

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8396 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 8337

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8336 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 8277

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8276 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 8217

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8216 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 8157

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8156 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 8097

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8096 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 8037

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8036 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 7977

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7976 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 7917

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7916 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 7857

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7856 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 7797

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7796 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 7737

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7736 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 7677

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7676 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 7617

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7616 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 7557

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
              ||||||||||||||||||||||||||
Db       7556 AAACAAGAAATTGAATCCCGATCTTAA 7530
```

```
RESULT 3
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             5.5%;  Score 61.4;  DB 3;  Length 1141;
  Best Local Similarity   11.5%;  Pred. No. 0.00015;
  Matches  108;  Conservative  340;  Mismatches  486;  Indels    3;  Gaps    2;

Qy         78 TTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTTAATGAGCAATACTATAG 137
              ::|:|:   : :::       ::|  :::  : |  | :::: ::::   |: :|:| :
Db         58 KWTWARMYCKYRRWYNNKSRWWKGWYKKKWYBCANNTSBRYHARRWKDMKTAYBMTMTNK 117

Qy        138 AGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTAAGAGCTC 197
              :| ||  : :: : |: :|::::::| |    : |:  :  ::::| ::: :     | :
Db        118 WGKTGWRHRYWRWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNATGWDD 177

Qy        198 TGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCGTGT 257
              :  :: :    | |:|:  : : : : : ::    :|:: :: : : ::  ::: ::::: :::
Db        178 DTKYHMWNNNGCBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRK 237

Qy        258 TCCTTCGGGAAATTTGAATAAAAAA--GCATGGATTGGTTATATTTATGACTTTCAACAC 315
              :   :|::: : :|:: |::::    :|: : ::::|:      :|: :
Db        238 VRRWVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRAKYWGW 297

Qy        316 TGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAA 375
               : ::      ::: ||:  |  :|  |:||:| ::  :  :   :   :       :
Db        298 NRABVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNWTY 357

Qy        376 TTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCA 435
              :: :::   :: : :  : ::|  |::  :::|  :: ::    |||:::::  ::  :
Db        358 KKARHBARWDWVWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKM 417

Qy        436 AATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCT 495
              :::: |      ||: :: :|  : ::  ::: |  |:      ::|       :  : |
Db        418 DMDWBGTYNNNNNGGRTYYGWTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKT 477

Qy        496 CAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGAT 555
              ::   :      ::  : |    ::   ::::|:     ::::: |  ::: :: ::::
Db        478 YWNNCYWKSMTNGKSHRBAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBVCSKWW 537

Qy        556 TATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTA-GGGC 614
               : : |   :   ::::::|  : :::::|: :|:   :  | ::|::: |:  : :
Db        538 NNYAAWYTKSSWNYTSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWS 597

Qy        615 ATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGCACGGGAGCTACTCA 674
              :|::: :|   ::::: : :|: : :    : :::|     :
Db        598 HTWBHBRAGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNNNNNATC 657

Qy        675 AGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCAAAAAAGCTCGGAAT 734
              |::::   :::: :   ::: ::: |      |:       :    : :: :
Db        658 ARDDYYAASRWYAMANAKWYYYKBAANNAYYTHANNWWGCWNNATDTRRTMWKNNNNNNA 717

Qy        735 TGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAAATTGAATTAATCAA 794
              :       |: |  |: : |  : |  ::  :|     : ::|:|  :  :::   |
Db        718 GTWKNNNNNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADAAABTTDKRNNGAY 777

Qy        795 AAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCTGGAGGGGGGGTAAC 854
              ||      : | :    |:  :|              | : ::: :    ||
Db        778 TKYTTTNNNNTYRGVVTNTAARDGWANNNNNNNNNNNNNNNGWSDMWVTWWAYANYGTNNN 837
```

```
Qy        855 ATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATAGATGTCAATAAAGA 914
                  : ::|  :::  ::::: |         : ::||:|:: |:  ::  :|:
Db        838 NNNNNNNAYAWWTNKWYYTTDDRWRBAYTNNNNNNRMAYYGAYADDYAYYMSDTCDAWMK 897

Qy        915 AGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTATTCATTAAATGACGC 974
              :: | :     : | ::|       |:  ::  :||:       :: | : :::
Db        898 WDATKMNNATTYNRGTAWRTNNNNNNMTMKTKYYBHAAWNNNNNNGKMCTAHTWWVCKAT 957

Qy        975 GATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAA 1011
              : |:| ::    | ::    ||:  : ||| | : ||
Db        958 KTTKGCWMNCTTTCRKYKNNCTWYTWMTTTRTTWYAA 994


RESULT 4
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

  Query Match             4.9%;  Score 54;  DB 3;  Length 832;
  Best Local Similarity   11.6%;  Pred. No. 0.0064;
  Matches   42;  Conservative  170;  Mismatches  150;  Indels    0;  Gaps    0;

Qy        553 GATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGG 612
              :::::::||::|  :|:  :    : :::   :  :::::::  :  : ::: :|::  :
Db          2 RWYWWKYTTWYAKCWTKWKWSWSYWMYWKWYYMKTYWRWRRKKKKAWWKYWKTWTWWYWR 61

Qy        613 GCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGCACGGGAGCTACT 672
              :: |       :|  ::  :  ::: :  ::  ::  : :|:||:|: ::: ::: :
Db         62 YAMWGTYKKKAMCRTKTKKKKKKGYMWMWYWGWRRSYMAMWTRTWTGYAYYRSMMYWWRY 121

Qy        673 CAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCAAAAAAGCTCGGA 732
              :: :  : : | ::::|  ::: :: |:  :::: |  :  |:  ::: : : :  ::
Db        122 RCWKKKAYYRKTTCYSSKGWTWWKRWKKAWTTWWWKKTYYWAATRYWWMMCWTKRWRASW 181

Qy        733 ATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAAATTGAATTAATC 792
              ::  ::   : : |:: ::    |::  :  : |:: |:::: :: :::: |:|| |
Db        182 WYCWWWGKARKWSTWRKSRSYASARSAKRCCYSCSWGAMSWKYMWRMWRWRGWATGAGMK 241

Qy        793 AAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCTGGAGGGGGGGTA 852
              |::|     :::  ::: : ::::|:  :::  :|:|:      : :||: |::|: :|
Db        242 AWRASCMMRRKYAGKSKTSYKSMWMCWTRSWKYCYTKARWTGYYCYRKGGMWGKRGRWYA 301

Qy        853 ACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATAGATGTCAATAAA 912
              : :::    :  :::  | :| :::::::  |::  : : | | | :  : :::|::
Db        302 SKKYMWKRWWWCWARMYRYSTGTRASMWWRRWYYTMMMKWWKYAWARAAWRWWAMWWAWR 361

Qy        913 GA 914
              :|
Db        362 RA 363


RESULT 5
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             4.8%;  Score 52.6;  DB 3;  Length 1141;
  Best Local Similarity   12.6%;  Pred. No. 0.014;
  Matches  100;  Conservative  274;  Mismatches  411;  Indels    9;  Gaps    2;

Qy        267 AAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAACACTGTTACTATCC 326
              ::: |:  ||  |::  :|::|||   :: :  | :   : |: :: ::|:  | :
Db       1104 RMDATKWSATGTAWWTNHAKRGATMCWYWYWTGTNRRWCMRTYAMRTWYTRSNANWSCAT 1045

Qy        327 TTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCAA 386
              :::::| :::| : :      ::  ::   ::      : ::  |   ::  | ::  : ||
Db       1044 KBMWWTMKWYATKYRTAWYAMWCAWRNNNMWCATNGYAKSCATNNAMWYATTRWAAYAAA 985

Qy        387 TTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATATGT 446
              ::       ::::   |  ::  :  :| : : : :|| ::|       :  :::: | :
Db        984 KWARWAGNNMRMYGAAAGNKWGCMAAMATMGBWWADTAGKMCNNNNNNWTTDVRRMAMKA 925

Qy        447 TGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATG 506
              :        :|    :||    :||:   :|: |: ::: | ::|: |   : ::
Db        924 KNNNNNNAYWTACYNRAATNNKMATHWMKWTHGAHSKRRTRHHTRTC--RRTKYNNNNNN 867

Qy        507 GTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAAT 566
              |:  : :  : ::   : :|          | : |:|: |::  ::
Db        866 ARTVYWYHHAARRWMNAWWTRTNNNNNNNNNNACRNTRTWWABWKHSWCNNNNNNNNNN 807

Qy        567 TTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTA 626
              : ::    ::  :|    || :  :|    ::    |||: : | :|::  :::::
Db        806 NNNTWCHYTTANABBCYRANNNNAAARMARTCNNYMHAAVTTTHTDWCYKTWMNTWYWDM 747

Qy        627 TACTGAATATAATCCTGATGTTTATTTAGTATGCACGGGAGCTACTCAAGATTATCGATT 686
              :  |   :| :| :|        :| |         :| : : |  :| ::   :|
Db        746 MTTMBTTTTRNMTTSTNMTNNNNNNMWACTNNNNNNMWKAYYAHATNNWGCWWNNTDA-- 689

Qy        687 CCCTGGATATTTTAATGAATTGATGGTTTTGGCAAAAAAGCTCGGAATTGAATCGAAAAT 746
                ::|  || :: :::|   |: : ||::::   |
Db        688 -----RRTNNTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNNNNNNNNNNNNNNSCCTC 634

Qy        747 TAAGATATTAGGGCATATACCTAAACTTGAACAAATTGAATTAATCAAAAATTGCATTGC 806
              |::   ::|::|: :|:   ::::: |   : ::|:  ::::| ::: :::|:    | |
Db        633 TRMMTMRWTMKGDGMTVRKKVKWRDTTCTYVDVWADSWVWWYANWMRCRDVTYTRNNTYC 574

Qy        807 TGTAATACAACCAACCTTATTTGAAGGCGGGCCTGGAGGGGGGGTAACATTTGACGCTAT 866
              :::|:    ::  |: : ::  : :::: :      :  :|::: ::   | : ::
Db        573 KSYAHSYWYWSNNAMWYRRYSARNWSSMARWTTRNNWWMSGBVRMRWAGTMWWRHWNNNN 514

Qy        867 TGCATTAGGGAAAAAAGTTATACTATCTGACATAGATGTCAATAAAGAAGTTAATTGCGG 926
              |: ::::  :::|:  | : :      :      :  :::: |  :|  :::|: :|
Db        513 TDTRYYWWWKRWARBTTTVYDSMCNAKSMWRGNNWRAMKMWWAANNDAGAMDHWTYWMGN 454

Qy        927 TGATGTATATTTCTTTCAGGCAAAAAACCATTATTCATTAAATGACGCGATGGTAAAAGC 986
              ::  : ::     :|:: :        |  ::::  ::: :: :   |       | :
Db        453 NTMWMRRAWKMMNMAWCRRAYCCNNNNNRACVWHKHKMWRWTWKYMWKAACNNNNBKAMY 394
```

```
Qy        987 TGATGAATCTAAAATTTTTTATGAACCTACAACTCTGATAGAATTGGGTCTCAAAAGACG 1046
Db        393 MRVAWMMYSRDTTNTDWMMWTSDWBWHWYTVDYTMMRAWNNNNNNNWRBCKTTSWMWWMD 334

Qy       1047 CAATGCGTGTGCAG 1060
Db        333 HMNTHCTYGNNTWG 320


RESULT 6
US-09-662-254B-23
; Sequence 23, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-23

  Query Match             4.7%;  Score 52;  DB 3;  Length 50000;
  Best Local Similarity   45.2%;  Pred. No. 0.041;
  Matches  190;  Conservative    0;  Mismatches  230;  Indels    0;  Gaps    0;

Qy        146 TTACACAAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTAAGAGCTCTGCCTTTT 205
Db      17297 TCACAATTATACACAAATAAAAGTTTTAATTCATATATAATATTAATATTTTTTTTTAAT 17356

Qy        206 ATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTCGG 265
Db      17357 ATTCTTATTTTATATTTCTTAGAATATATATTATTCTTTATTTTATTTATGATATATTTA 17416

Qy        266 GAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAACACTGTTACTATC 325
Db      17417 GAATATTTTATAAAATATTTACTATATGTTATATTGTCTTTTATATACATATATTGTTTA 17476

Qy        326 CTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCA 385
Db      17477 TTGTTATATTTATATTGCAATGTAGACATTTCCAATATATTTTTTATATATTTATCCATT 17536

Qy        386 ATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATATG 445
Db      17537 TTTATATATTAATTTAAAATTTCAAAAACATTTATTTGTCTAATGTTATTAATAATTTTG 17596

Qy        446 TTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAAT 505
Db      17597 TATATTTATTATATAACAAATATAAATATTCTGAAGATATAAAGTGCATTGAATTTGCGT 17656

Qy        506 GGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAA 565
Db      17657 TGTAATCTAATTTTTTATATTTTTTTTTAATTTCTATATCTGGATAAAATTTTTTTATAA 17716


RESULT 7
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             4.6%;  Score 50.8;  DB 3;  Length 640681;
  Best Local Similarity   47.7%;  Pred. No. 0.13;
  Matches  148;  Conservative    0;  Mismatches  162;  Indels    0;  Gaps    0;

Qy        522 TGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTG 581
Db     327211 TGTTGCTGTTGCTAGAGCATTTATTAATAAGCCATCTTTAATAATAGCAGATGAACCTAC 327270

Qy        582 GAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCC 641
Db     327271 TGGAAATTTAGATGAAGACAATACAAATATTATCTTTAATTTAATCACAGAATTAAATTC 327330

Qy        642 TGATGTTTATTTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAA 701
Db     327331 TGATTATAATACTTCGTTCATAATTGCCACACATGATCCTACTTTAATAAAAAAGATACC 327390

Qy        702 TGAATTGATGGTTTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCA 761
Db     327391 TGTTTTGTTTAAAATAGAAAACAATCAGATATTTAATTATGAAAGTTAATTAAAAAGAAA 327450

Qy        762 TATACCTAAACTTGAACAAATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAAC 821
Db     327451 CATTCAATGAATTCTTTACCTTTTTTAATTGCAAAAAAATTTTCTCTTAAGAAAAACAAA 327510

Qy        822 CTTATTTGAA 831
Db     327511 AATTATTTAA 327520


RESULT 8
US-09-662-254B-33
; Sequence 33, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
;   LENGTH: 1410
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1)..(1410)
;   OTHER INFORMATION:
```

US-09-662-254B-33

Query Match 4.5%; Score 49.4; DB 3; Length 1410;
Best Local Similarity 46.0%; Pred. No. 0.077;
Matches 167; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

```
Qy         11 CGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAAAGTATATTAA 70
Db        974 CGGATAATACAAAATGTTTAATTTTGACAGATTATAGATTACACATGATGAATATATATA 1033

Qy         71 AAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTTAATGAGCAAT 130
Db       1034 ATTTATTAAAAAAAACACACTTACAAAATATAATTTATATATATGATGTAAAAAATAAAA 1093

Qy        131 ACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTA 190
Db       1094 AATGTAATGATTTGTTAACAGAAATTAAAAATAAGAATGAAAAATTTATTATTATATCAA 1153

Qy        191 AGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCA 250
Db       1154 CTATATCTGCTTGTTCTGAATCATTAGATATTAATAATTTAAATACTTTTCATGTTTTAT 1213

Qy        251 TGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTC 310
Db       1214 TACCTATTACTAATTCTAAAACAATAAAACAATGCATAGGTAGAATTATGAGAAATATGA 1273

Qy        311 AACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTT 370
Db       1274 ACGAAGATAAATATACTTATATATATAATTTTTCTAACATCAATAACATGATTAATATGT 1333

Qy        371 TTA 373
Db       1334 ATA 1336
```

RESULT 9
US-09-662-254B-24
; Sequence 24, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-24

Query Match 4.5%; Score 49.4; DB 3; Length 50000;
Best Local Similarity 46.0%; Pred. No. 0.16;
Matches 167; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

```
Qy         11 CGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAAAGTATATTAA 70
Db      10334 CGGATAATACAAAATGTTTAATTTTGACAGATTATAGATTACACATGATGAATATATATA 10393

Qy         71 AAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTTAATGAGCAAT 130
Db      10394 ATTTATTAAAAAAAACACACTTACAAAATATAATTTATATATATGATGTAAAAAATAAAA 10453

Qy        131 ACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTA 190
Db      10454 AATGTAATGATTTGTTAACAGAAATTAAAAATAAGAATGAAAAATTTATTATTATATCAA 10513

Qy        191 AGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCA 250
Db      10514 CTATATCTGCTTGTTCTGAATCATTAGATATTAATAATTTAAATACTTTTCATGTTTTAT 10573

Qy        251 TGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTC 310
Db      10574 TACCTATTACTAATTCTAAAACAATAAAACAATGCATAGGTAGAATTATGAGAAATATGA 10633

Qy        311 AACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTT 370
Db      10634 ACGAAGATAAATATACTTATATATATAATTTTTCTAACATCAATAACATGATTAATATGT 10693

Qy        371 TTA 373
Db      10694 ATA 10696
```

RESULT 10
US-09-949-016-17061
; Sequence 17061, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17061
; LENGTH: 96922
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17061

Query Match 4.3%; Score 47.2; DB 3; Length 96922;
Best Local Similarity 44.4%; Pred. No. 0.56;
Matches 276; Conservative 0; Mismatches 343; Indels 3; Gaps 2;

```
Qy        195 CTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCG 254
Db      38321 CTTTCAGTTTTATTTATCTGTTTTTGCCATACATTATATATTTATATATATATTTATATA 38380

Qy        255 TGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAACA 314
Db      38381 TATATTTAATATATATAAATATATTTTTATATATATTTGATATATAAATATATTTATATA 38440

Qy        315 CTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAA 374
Db      38441 TATATTTGATATATAAATATGTTTTTATATATATTTGATATATAAATATGTTTATATATA 38500

Qy        375 ATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGC 434
Db      38501 TTTGATATATAAATATGTTTATATATATTTGATATATAAATGTTTATATATATTTGATAT 38560

Qy        435 AAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCC 494
Db      38561 ATAAATATATGTTTATATATATTTGATATATAAATATATGTTTATATATATTTGATATAT 38620

Qy        495 TCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGA 554
Db      38621 AAATATATGTTTATATATTTAATA-TATAAATATATGTTTATATATTTAATATATAAATA 38679
```

```
Qy        555 TTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGC 614
              |  |||||||  ||   ||  || ||    |   |||     ||           | |   |
Db      38680 TATTTTTATATATTTATATATATATTTATATATTAATATATATATTTAATAGATAAATAC 38739

Qy        615 ATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGCACGGGAGCTACTCA 674
              || || | || |||   ||    ||   || |||| |   ||   ||       |
Db      38740 ATATATATTTAATAGATAAATACAT--ATATATTTAATAGATAAATACATTTATATATAT 38797

Qy        675 AGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCAAAAAAGCTCGGAAT 734
              |   |||  |||   |  |||  |  ||  ||  ||  | |     |   ||      |||
Db      38798 AATATATATATTTAATATATAAATACATTTATATATAATATATATATTTAATAGATAAAT 38857

Qy        735 TGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAAATTGAATTAATCAA 794
              | |     ||| | |||||     ||||   || |  |  | | ||  |||   |  |
Db      38858 ATATTTATCTATTTACATATTTATATATATTTATATATATATAAATATAAAATATTTATA 38917

Qy        795 AAATTGCATTGCTGTAATACAA 816
              |  |  |    | || || ||
Db      38918 TATATTTACATGTATATTATAA 38939
```

RESULT 11
US-08-998-416-288
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241RP
US-08-998-416-288

Query Match 4.2%; Score 47; DB 3; Length 837;
Best Local Similarity 44.0%; Pred. No. 0.24;
Matches 244; Conservative 0; Mismatches 310; Indels 1; Gaps 1;

```
Qy        103 ATTTCATTAAATAGATTTAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATA 162
              |||| ||| |  |   |||||    | |  | ||         ||||   |  | | |||
Db         17 ATTTAATTTATAAAGATTAATATAAACTTTTTTATTATAATATTTAAGTATTAAATTATT 76

Qy        163 GAGACGAATCTTACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGT 222
              | || | | ||| | ||||| ||   |  |  | |    ||  |     || | |||
Db         77 TAAACTATTATTATCATTATTTAATAAATTAATTATTTGATTATTAATACTTATTATATA 136

Qy        223 AGCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAA 282
              |     |   |  ||| |         |||   |  |  ||    || | | || | || |
Db        137 ATTATTATATAATTTACTTAATTCATCATTATTAATATTTATATAATTATAAAAATAATA 196

Qy        283 GCATGGATTGGTTATATTTATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAG 342
                  ||   |  ||||||     |     | | |||   |  ||  |    || ||
Db        197 TTTAATATGAATACTATTTAGTCTATGTTCAAATTTTAAATTAGTTATTAAAATATTATT 256

Qy        343 CGAGAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATT 402
              || |     ||        |     || |||||| |||  |||||       | | ||||
Db        257 AGATATTATTATTTTCTTTAATAAATTATTAAATAGATTATCAATAATTAATATATTATT 316

Qy        403 ATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCA 462
              | | | ||  ||| |  || ||     |     || |  | | |   | |||||
Db        317 -TATTAATTGTTTATTAAAATAATATATTTTATTATTATAAAGATTTAATTTATTTAAAT 375

Qy        463 AAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCT 522
              |   || ||| |  |   | | |     |  |   |  |||| |  | ||||| | |
Db        376 ATTGTAAATTATTATTTTTATTATAATATCTATTTTTATAAATATTATGTTGATTTATAT 435

Qy        523 GGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGG 582
              |   ||       ||| | |        | |   || ||| || | |   | ||
Db        436 TATTTAATCTTTTTATAAGAATTATTATTAAAATTAATTTTAACTTTAATTTCTTATTAT 495

Qy        583 AAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCT 642
              || | | || ||  ||    || ||   ||||| |  | | || | | |  | |
Db        496 TAATTTTTATATTATTTAATAAATTATATTTCATTTTATTTATTTATTTATTTAATTAAA 555

Qy        643 GATGTTTATTTAGTA 657
                ||  || | ||
Db        556 TTAATTATTTAATTA 570
```

RESULT 12
US-09-662-254B-27/c
; Sequence 27, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 32392
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-27

Query Match 4.2%; Score 47; DB 3; Length 32392;
Best Local Similarity 45.6%; Pred. No. 0.5;
Matches 202; Conservative 0; Mismatches 240; Indels 1; Gaps 1;

```
Qy        348 AATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGT 407
              ||  |||  ||  | | |   | |    || |     ||      ||| ||| |   | |
Db      28481 AACTGATTTAAATATTTTACCTATAGGTTTAAGAAAAAAAATAAATAAAAATGTGTGTTT 28422

Qy        408 TAATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACT 467
              ||||    | |   |  ||    ||   ||| ||    || | ||||||| |  | | |
Db      28421 TAATTGCAAATGTATGTTTTTTAATAATAATCAAATAATTTGTAATTATTGTAAATTATT 28362

Qy        468 ACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAA 527
              |  ||||         ||    ||    | | |||                ||   |     |
Db      28361 ATGTTCTGGATGTAATAAATTATATAATAATTTATCGATTAAAAAATTTTCTACAAAATA 28302

Qy        528 TATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACA 587
              |  |   | |||| | | | ||  | || |  || | ||  ||  || |    ||| |
Db      28301 TGGTTATAGATATGAAAATAAAAAATTTAATATATTATTATGTGATATATGTAAAAAAAA 28242

Qy        588 TAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGT 647
              || |||        | ||  | |     |||   | || ||  | |  ||| |  | ||||
Db      28241 TACAGAAATATGTATTGAGTGTCATAAATTATTATTTAATTATAATAATATTGATTATGT 28182

Qy        648 TTATTTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCC-TGGATATTTTAATGAAT 706
              | |||  |   |  |    || |||||||  |          || | |||||   | |||
Db      28181 AGAATTAAGAAATATAGATAGTATTCAAGATAAAGTCGGAGTATGTAAATTTTGTTTAAT 28122

Qy        707 TGATGGTTTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATAC 766
              | ||  |||  |  |  ||  |  |  |   |   || | |  |||| |   ||||||
Db      28121 TAATATTTTATGTGATGAATGTAATAGATATTTAACAACTAATTATATAAATTCATATAA 28062

Qy        767 CTAAACTTGAACAAATTGAATTA 789
                |||  |      |  | | |||
Db      28061 TAAAAATAATTATACGTTATTTA 28039
```

RESULT 13
US-09-601-198-43
; Sequence 43, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellson Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; TITLE OF INVENTION: UREALYTICUM
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 3399
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-43

Query Match 4.2%; Score 46.8; DB 3; Length 3399;
Best Local Similarity 47.9%; Pred. No. 0.35;
Matches 135; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

```
Qy        369 TTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTAC 428
              |||  ||| | |  ||      |   ||| ||| |||| | | |    ||   | |  |||
Db       1806 TTTAAAAATAAAACTAGTGGGTAATAAAAATCTTATAGATGAAAATCAAAAATTATCTAC 1865

Qy        429 CGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCC 488
                  ||| | |||||||  || | |  |  |||| |||   | || |   || |
Db       1866 AATGATTAATGATTATGTTGAAAACTTTGATATAAAAAAACACCTTGTTGCTAATAATAA 1925

Qy        489 ATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGA 548
```

```
              | | | |||| ||     || |  |||||   |  ||||  |       |||
Db       1926 AAACTCCCAATCAATTGTTTTTGACGATTATGTTATAAATAAAGTATTTAGTAAAGGAAT 1985

Qy        549 CAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTT 608
               ||   |  ||||  || |   | | ||        ||  |||| |   || | |   ||
Db       1986 AAACAGTGTTTTTGCAAATATTATTAAACAAGAACAAATTAAAAAAAGTGTAGCAAAATT 2045

Qy        609 TAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTA 650
                | ||||||   || || |   | |||  |  | || |
Db       2046 CTTTGAATTTAATGATTTAATTTTAAAAATAATTTTCTTGA 2087
```

RESULT 14
US-09-662-254B-23/c
; Sequence 23, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-23

Query Match 4.1%; Score 45.6; DB 3; Length 50000;
Best Local Similarity 47.3%; Pred. No. 1.1;
Matches 138; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

```
Qy        367 TTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATT 426
              | || |||||  ||  | |||       || |||| |||   |  |    | |    ||
Db      17421 TATTCTAAATATATCATAAATAAAATAAAGAATAATATATATTCTAAGAAATATAAAATA 17362

Qy        427 ACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGT 486
              |     ||| ||| || ||  | |   | || |  |  || ||   ||   | || |
Db      17361 AGAATATTAAAAAAAAATATTAATATTATATATGAATTAAAACTTTTATTTGTGTATAAT 17302

Qy        487 CCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATT 546
                  |   || | ||  |  |||||||   || | ||  |    ||| |||||||
Db      17301 TGTGATAAACTCATAGATTGTATATCTGATTATGATGATTATGTCAATAAAAATAATATA 17242

Qy        547 GACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCT 606
                | | || ||| ||||| |   | |       || ||  | ||| |  ||
Db      17241 ATTATGAATGATTATATAAATATGATTATTAACTGTAATAAAAAACAAAATATTGAATTA 17182

Qy        607 TTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTAT 658
               |||     ||||   | ||  |||  | | |    ||  ||||    |||
Db      17181 ATTAAAATTATTAATGATAATTTTGATAAAATTATATATAATTATGATATAT 17130
```

RESULT 15
US-09-662-254B-24/c
; Sequence 24, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1

```
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-24

  Query Match             4.1%;  Score 45.2;  DB 3;  Length 50000;
  Best Local Similarity   49.0%;  Pred. No. 1.4;
  Matches  149;  Conservative    0;  Mismatches  153;  Indels    2;  Gaps    1;

Qy        336 TAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAA 395
              ||  ||    ||||||    |     || |  ||   ||    ||   ||      |||
Db      36169 TAAGAACGTCGAAATAATAGATTTTGATTTAATTGAAAATGAAATAGACATGAAACCTAT 36110

Qy        396 CAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAATTA 455
              ||||||||  | ||   |  |  || |  || ||  |  |||||  ||      || ||
Db      36109 TCATATTATTATAAAAAAAATTAGAGATGATAACGGAGATAATAAGCATAAACATAAATA 36050

Qy        456 TTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGA 515
              |  |  ||  |   ||  |  | | |||| || ||   ||| || |    |   | |
Db      36049 TGTTTTAATTCCGAATATTAAT--AATAAGTCAATACCTATAATAAATAAAAATTAATTA 35992

Qy        516 TTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCAATCA 575
              |||   |  | |||||  ||| ||| | | | |     | || ||   ||| |  ||  |
Db      35991 TTATATTTATTATATTTACAAGTATTAAAATTATTTTCAATACTTGTAAATGTAAAAATA 35932

Qy        576 ATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATA 635
              ||  |  ||| | ||| |  ||     |  |  |  | |||  || ||||  | |   |
Db      35931 ATAATTTAAAAAAAAAAAATATAATCTTTTTGATGTGCCATAAAATATTTGAAATTTGTT 35872

Qy        636 TAAT 639
              || |
Db      35871 TATT 35868


Search completed: May 11, 2006, 05:22:57
Job time : 157.411 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:01:11 ; Search time 730.653 Seconds
(without alignments)
12528.801 Million cell updates/sec

Title: US-09-423-093B-2_COPY_5365_6471
Perfect score: 1107
Sequence: 1 attcttttaccgagaaatga.....aaattgaatcccgatcttaa 1107

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1107	100.0	14024	3	US-09-423-093A-2	Sequence 2, Appli
c	2	1107	100.0	14187	5	US-10-114-170-121	Sequence 121, App
c	3	64.8	5.9	3089	9	US-10-954-778-160	Sequence 160, App
c	4	64.8	5.9	4985	5	US-10-094-240-10	Sequence 10, Appl
c	5	64.8	5.9	4985	6	US-10-056-405-10	Sequence 10, Appl
c	6	61.8	5.6	3673778	6	US-10-312-841-1	Sequence 1, Appli
c	7	59.2	5.3	8056	8	US-10-473-126-386	Sequence 386, App
c	8	56.6	5.1	8056	8	US-10-473-126-240	Sequence 240, App
	9	56.2	5.1	16217	6	US-10-311-455-598	Sequence 598, App
	10	56.2	5.1	335913	3	US-09-754-853A-2	Sequence 2, Appli
	11	56.2	5.1	335913	3	US-09-754-853A-3	Sequence 3, Appli
	12	55.4	5.0	11790	6	US-10-311-455-516	Sequence 516, App
	13	54.2	4.9	5487	6	US-10-311-455-1571	Sequence 1571, Ap
	14	54	4.9	6132	6	US-10-311-455-836	Sequence 836, App
c	15	53.4	4.8	11143	10	US-11-097-143-16492	Sequence 16492, A
	16	52	4.7	50000	8	US-10-706-635-23	Sequence 23, Appl
	17	51.4	4.6	34688	7	US-10-433-793-90	Sequence 90, Appl
	18	51.2	4.6	8056	8	US-10-473-126-386	Sequence 386, App
	19	50.8	4.6	21537	6	US-10-311-455-1972	Sequence 1972, Ap
	20	50.8	4.6	640681	3	US-09-790-988-1	Sequence 1, Appli
	21	50.6	4.6	7025	7	US-10-257-166-141	Sequence 141, App
	22	50.6	4.6	7025	7	US-10-240-454-45	Sequence 45, Appl
	23	50.6	4.6	3673778	6	US-10-312-841-1	Sequence 1, Appli
	24	49.6	4.5	543	7	US-10-424-599-111337	Sequence 111337,
	25	49.4	4.5	1410	8	US-10-706-635-33	Sequence 33, Appl
	26	49.4	4.5	8961	6	US-10-240-453-302	Sequence 302, App
	27	49.4	4.5	50000	8	US-10-706-635-24	Sequence 24, Appl
	28	49.4	4.5	70768	5	US-10-135-322-13	Sequence 13, Appl
	29	49	4.4	11143	10	US-11-097-143-16492	Sequence 16492, A
	30	48.8	4.4	960	5	US-10-198-846-6381	Sequence 6381, Ap
	31	48.8	4.4	61020	7	US-10-221-714A-514	Sequence 514, App
	32	48.6	4.4	5398	6	US-10-311-455-115	Sequence 115, App
	33	48.4	4.4	2310	7	US-10-221-714A-460	Sequence 460, App
	34	48.4	4.4	6636	6	US-10-311-455-764	Sequence 764, App
	35	48.4	4.4	8056	8	US-10-473-126-240	Sequence 240, App
	36	48.2	4.4	6954	6	US-10-311-455-1364	Sequence 1364, Ap
	37	47.8	4.3	3673778	6	US-10-312-841-2	Sequence 2, Appli
	38	47.4	4.3	6963	6	US-10-311-455-951	Sequence 951, App
c	39	47	4.2	1517	3	US-09-938-842A-3281	Sequence 3281, Ap
c	40	47	4.2	1517	3	US-09-938-842A-3281	Sequence 3281, Ap
	41	47	4.2	6464	6	US-10-311-455-487	Sequence 487, App
	42	47	4.2	17421	5	US-10-239-676-54	Sequence 54, Appl
	43	47	4.2	17421	6	US-10-240-453-56	Sequence 56, Appl
c	44	47	4.2	32392	8	US-10-706-635-27	Sequence 27, Appl
	45	46.8	4.2	3399	6	US-10-349-680-137	Sequence 137, App

ALIGNMENTS

RESULT 1
US-09-423-093A-2
; Sequence 2, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 14024
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-423-093A-2

Query Match 100.0%; Score 1107; DB 3; Length 14024;
Best Local Similarity 100.0%; Pred. No. 1.1e-215;
Matches 1107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5365 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 5424

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5425 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 5484

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5485 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 5544

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5545 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 5604

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5605 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 5664
```

```
Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5665 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 5724

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5725 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 5784

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5785 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 5844

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5845 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 5904

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5905 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 5964

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5965 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 6024

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6025 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 6084

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6085 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 6144

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6145 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 6204

Qy        841 GGAGGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6205 GGAGGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 6264

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6265 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 6324

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6325 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 6384

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6385 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 6444

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
              ||||||||||||||||||||||||||
Db       6445 AAACAAGAAATTGAATCCCGATCTTAA 6471


RESULT 2
US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Publication No. US20030023075A1
;   GENERAL INFORMATION:
;        APPLICANT: Blattner, Frederick R.
;                   Burland,     Valerie
;                   Perna,  Nicole T.
;                   Plunkett, Guy
;                   Welch,  Rod
;        TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
;        NUMBER OF SEQUENCES: 265
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Quarles & Brady
;             STREET: 1 South Pinckney Street
;             CITY: Madison
;             STATE: WI
;             COUNTRY: US
;             ZIP: 53701-2113
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Word Perfect 8.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/114,170
;             FILING DATE: 01-Apr-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/453,702
;             FILING DATE: 03-DEC-1999
;             APPLICATION NUMBER: 60/110,955
;             FILING DATE: 04-DEC-1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seay, Nicholas J.
;             REGISTRATION NUMBER: 27386
;             REFERENCE/DOCKET NUMBER: 960296.95017
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (608) 251-5000
;             TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 121:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14187
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA  (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121

  Query Match             100.0%;  Score 1107;  DB 5;  Length 14187;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-215;
  Matches 1107;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8636 ATTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAA 8577

Qy         61 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8576 AGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTT 8517

Qy        121 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8516 AATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTT 8457

Qy        181 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8456 ATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATT 8397

Qy        241 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8396 CTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATT 8337

Qy        301 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8336 TATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGG 8277

Qy        361 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8276 AATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCA 8217

Qy        421 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8216 GTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCA 8157

Qy        481 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       8156 TTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATAT 8097

Qy        541 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8096 AATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCA 8037

Qy        601 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8036 ACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATGC 7977

Qy        661 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7976 ACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGCA 7917

Qy        721 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7916 AAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACAA 7857

Qy        781 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7856 ATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAAGGCGGGCCT 7797

Qy        841 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7796 GGAGGGGGGGTAACATTTGACGCTATTGCATTAGGGAAAAAAGTTATACTATCTGACATA 7737

Qy        901 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7736 GATGTCAATAAAGAAGTTAATTGCGGTGATGTATATTTCTTTCAGGCAAAAAACCATTAT 7677

Qy        961 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7676 TCATTAAATGACGCGATGGTAAAAGCTGATGAATCTAAAATTTTTTATGAACCTACAACT 7617

Qy       1021 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7616 CTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGTGCAGATTTTCTTTTAGATGTTGTG 7557

Qy       1081 AAACAAGAAATTGAATCCCGATCTTAA 1107
              ||||||||||||||||||||||||||
Db       7556 AAACAAGAAATTGAATCCCGATCTTAA 7530
```

RESULT 3
US-10-954-778-160/c
; Sequence 160, Application US/10954778
; Publication No. US20050153368A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: METHOD OF IDENTIFYING CHEMICAL AGENTS WHICH STIMULATE
; TITLE OF INVENTION: ODORANT RECEPTORS OF SENSORY NEURONS
; FILE REFERENCE: N1125
; CURRENT APPLICATION NUMBER: US/10/954,778
; CURRENT FILING DATE: 2004-09-30
; PRIOR APPLICATION NUMBER: 10/056,405
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 160
; LENGTH: 3089
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-954-778-160

Query Match 5.9%; Score 64.8; DB 9; Length 3089;
Best Local Similarity 46.1%; Pred. No. 0.0055;
Matches 286; Conservative 0; Mismatches 332; Indels 2; Gaps 2;

```
Qy         32 TTATAAGAGAAAAAGCATTTCCTTTTAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAA 91
              | |||| |  |  |  | |     | | ||  | | ||  |  |||  |||  |    ||
Db       2495 TAATAACAATAGCAATAATAAGAATAATAATAACAATAGCAATAATAATAATAATAATAA 2436

Qy         92 GGCCTCGATGGATTTCATTAAATAGATTTAATGAGCAATACTATAGAGATGCCTTTACAC 151
                  |   ||   | |||  | | | ||  |  |  | ||||   |      ||
Db       2435 TAATAATAATAATAATAATAATAATAATAAAAAAAAAATAATTATAATAAGAATAGTAATA 2376

Qy        152 AAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCAT 211
              | |||||||  | |  ||   ||  ||| ||| | ||| |    |      | ||
Db       2375 ATAATAATAATAATATTAAAAATAATATTAATAATAATAATAATAATAATAATAAGA 2316

Qy        212 ATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATT 271
              ||  | ||| ||   ||  | ||  || | |  |  |    |  || |     || |
Db       2315 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATCATAATAAAATTA 2256

Qy        272 TGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAACACTGTTACTATCCTTCAT 331
              | | ||  | |  | |  | || ||  | |   || | | | | ||  |  ||
Db       2255 TTATTATGATAATAATGTATAATAATGAATAACAATAATAATAAT-TAATAATAATAAAT 2197

Qy        332 TTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCG 391
              |  ||| | |   | |||||     ||  | | |  || ||| | | ||  | |||
Db       2196 TAATTAATTAATAATAAATAATAATAATTATTATTGTTATAATAATAAT-AATAATA 2138

Qy        392 CTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGA 451
              |||  |||| || |  ||||    |||    || ||   ||   |||||  ||  |   |
Db       2137 ATAATAATAATAATAATAATCATAATTATAATAATAATAATAATAATAATAATAATA 2078

Qy        452 ATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCG 511
              || ||  |   ||      ||  |  |       |||         | |   ||  |
Db       2077 ATAATAATAATAATAATAATAATAATAATAAAAGTAATAATTATAATAATAATAGTAATA 2018

Qy        512 CTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCA 571
              | || |   |  ||||| |   ||  |||||| | | ||  || ||  ||  | |   |
Db       2017 ATAATAATAATAATAATAATATTAATAATAATAATAATAATAATAATAATAATAATA 1958

Qy        572 ATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTG 631
              || | | |   ||  ||||  || ||   | |  |  ||       ||| | | ||| |
Db       1957 ATAATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1898

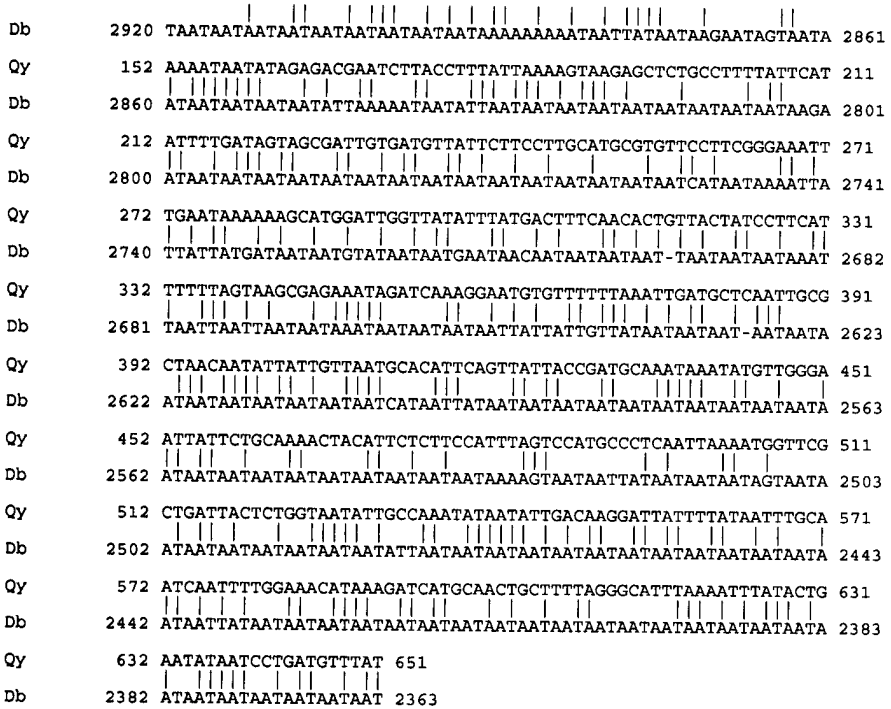
Qy        632 AATATAATCCTGATGTTTAT 651
              |  |||||  | ||  | ||
Db       1897 ATAATAATAATAATAATAAT 1878
```

RESULT 4
US-10-094-240-10/c
; Sequence 10, Application US/10094240
; Publication No. US20030082637A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
; FILE REFERENCE: N8289
; CURRENT APPLICATION NUMBER: US/10/094,240
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 10/056,405
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-094-240-10

Query Match 5.9%; Score 64.8; DB 5; Length 4985;
Best Local Similarity 46.1%; Pred. No. 0.0066;
Matches 286; Conservative 0; Mismatches 332; Indels 2; Gaps 2;

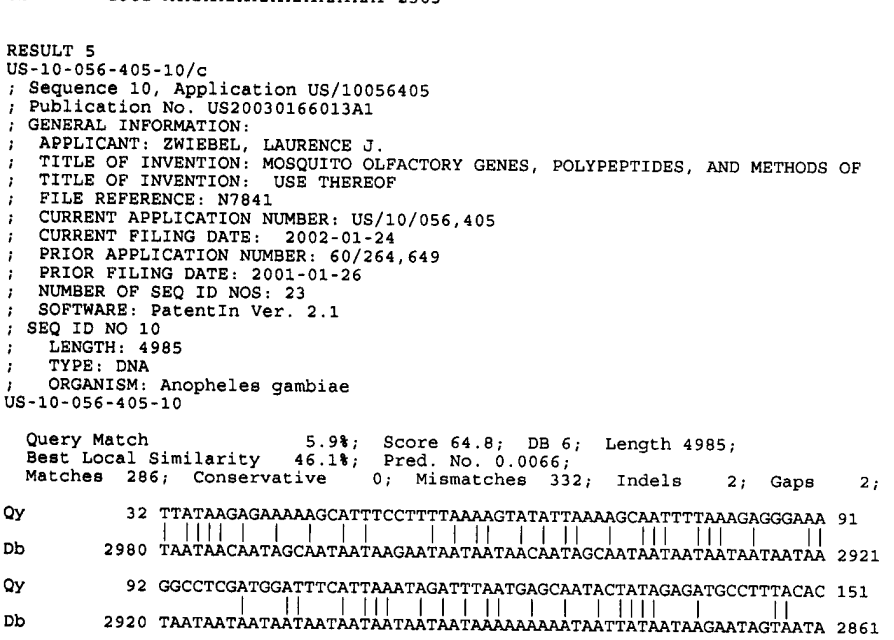
```
Qy         32 TTATAAGAGAAAAAGCATTTCCTTTTAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAA 91
              | |||| |  |  |  | |     | | ||  | | ||  |  |||  |||  |    ||
Db       2980 TAATAACAATAGCAATAATAAGAATAATAATAACAATAGCAATAATAATAATAATAATAA 2921

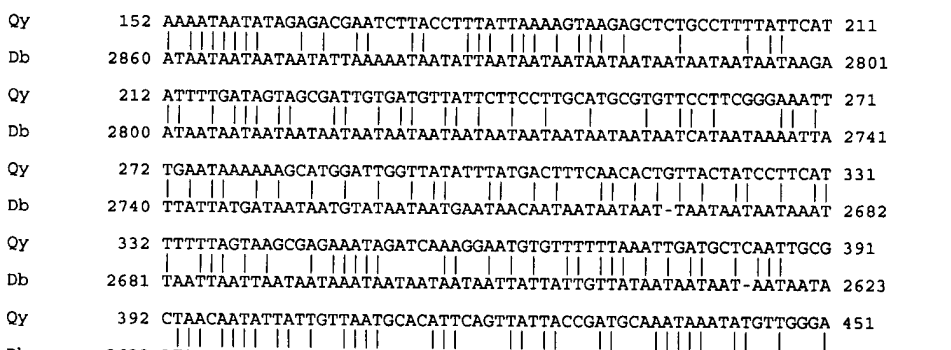
Qy         92 GGCCTCGATGGATTTCATTAAATAGATTTAATGAGCAATACTATAGAGATGCCTTTACAC 151
```



RESULT 5
US-10-056-405-10/c
; Sequence 10, Application US/10056405
; Publication No. US20030166013A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: MOSQUITO OLFACTORY GENES, POLYPEPTIDES, AND METHODS OF
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: N7841
; CURRENT APPLICATION NUMBER: US/10/056,405
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-056-405-10

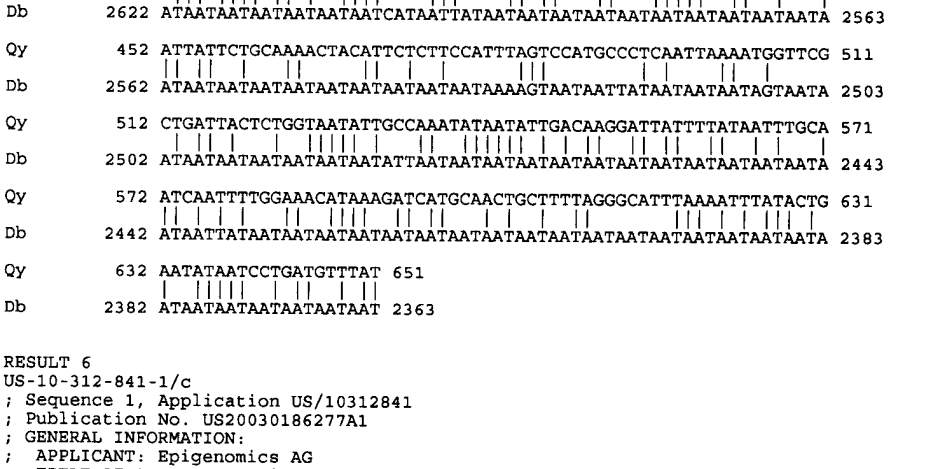
Query Match 5.9%; Score 64.8; DB 6; Length 4985;
Best Local Similarity 46.1%; Pred. No. 0.0066;
Matches 286; Conservative 0; Mismatches 332; Indels 2; Gaps 2;

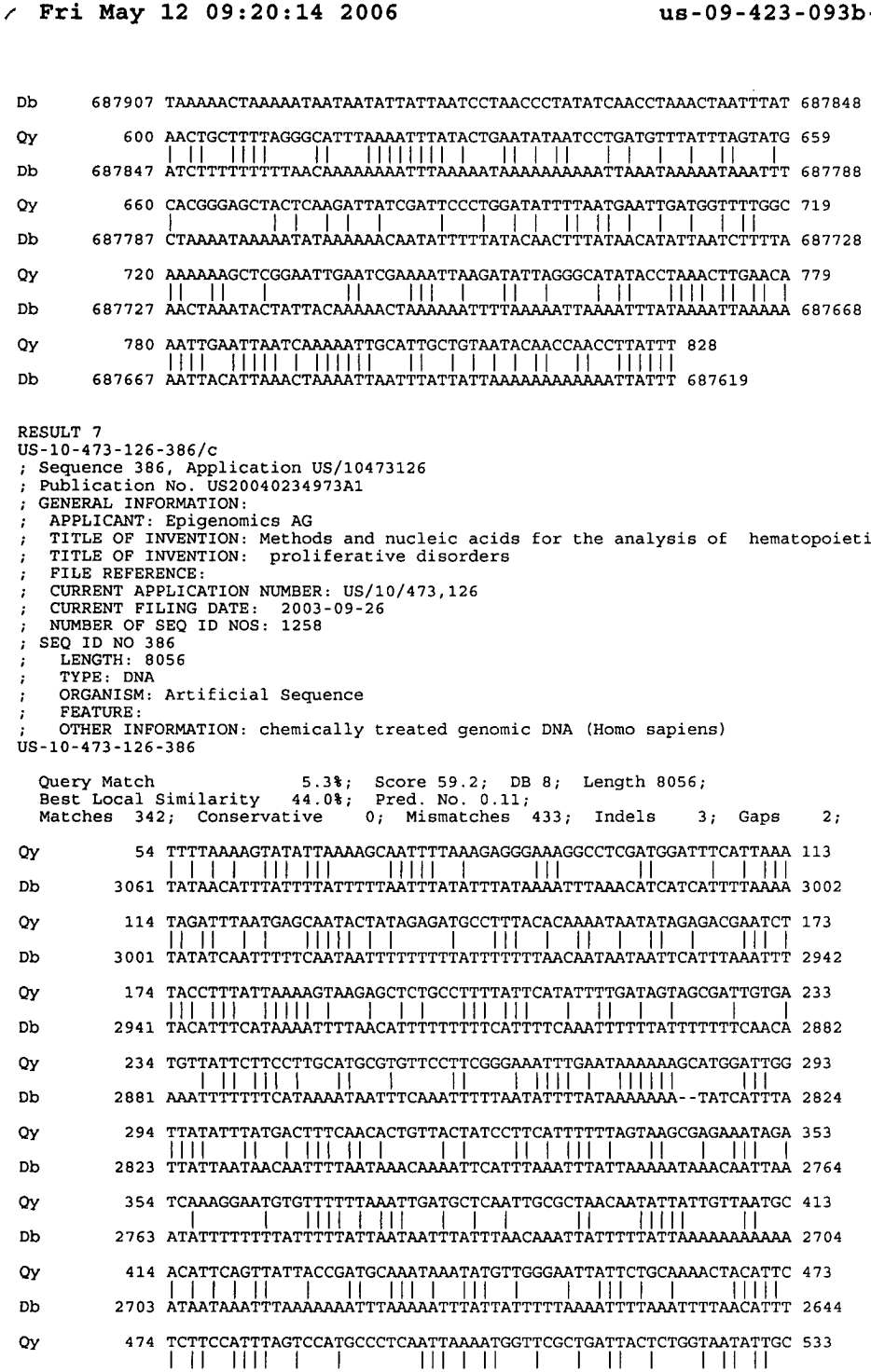




RESULT 6
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 5.6%; Score 61.8; DB 6; Length 3673778;
Best Local Similarity 46.9%; Pred. No. 0.33;
Matches 192; Conservative 0; Mismatches 217; Indels 0; Gaps 0;





```
Db     687907 TAAAAACTAAAAATAATAATATTATTAATCCTAACCCTATATCAACCTAAACTAATTTAT 687848

Qy        600 AACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTTAGTATG 659

Db     687847 ATCTTTTTTTTTAACAAAAAAAATTTAAAAATAAAAAAAAAATTAAATAAAAATAAATTT 687788

Qy        660 CACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGTTTTGGC 719

Db     687787 CTAAAATAAAAATATAAAAAACAATATTTTTATACAACTTTATAACATATTAATCTTTTA 687728

Qy        720 AAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACTTGAACA 779

Db     687727 AACTAAATACTATTACAAAAACTAAAAAATTTTAAAAATTAAAATTTATAAAATTAAAAA 687668

Qy        780 AATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTT 828

Db     687667 AATTACATTAAACTAAAATTAATTTATTATTAAAAAAAAAAAATTATTT 687619


RESULT 7
US-10-473-126-386/c
; Sequence 386, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Methods and nucleic acids for the analysis of  hematopoietic cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/473,126
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 386
;   LENGTH: 8056
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-386

  Query Match             5.3%;  Score 59.2;  DB 8;  Length 8056;
  Best Local Similarity   44.0%;  Pred. No. 0.11;
  Matches  342;  Conservative    0;  Mismatches  433;  Indels    3;  Gaps    2;

Qy         54 TTTTAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAA 113

Db       3061 TATAACATTTATTTTATTTTTAATTTATATTTATAAAATTTAAACATCATCATTTTAAAA 3002

Qy        114 TAGATTTAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCT 173

Db       3001 TATATCAATTTTTCAATAATTTTTTTTTATTTTTTAACAATAATAATTCATTTAAATTT 2942

Qy        174 TACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGA 233

Db       2941 TACATTTCATAAAATTTTAACATTTTTTTTTCATTTTCAAATTTTTTATTTTTTTCAACA 2882

Qy        234 TGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGG 293

Db       2881 AAATTTTTTTCATAAAATAATTTCAAATTTTTAATATTTTATAAAAAAA--TATCATTTA 2824

Qy        294 TTATATTTATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGA 353

Db       2823 TTATTAATAACAATTTTAATAAACAAAATTCATTTAAATTTATTAAAAATAAACAATTAA 2764

Qy        354 TCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGC 413

Db       2763 ATATTTTTTTTATTTTTATTAATAATTTATTTAACAAATTATTTTTATTAAAAAAAAAAA 2704

Qy        414 ACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTC 473

Db       2703 ATAATAAATTTAAAAAAATTTAAAAATTTATTATTTTTAAAATTTTAAATTTTAACATTT 2644

Qy        474 TCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGC 533

Db       2643 TTTTTTATTTTTTTTTTTTAAATTATTTATATTTATAATTTTTTTTTTCATTTATTTTTT 2584

Qy        534 CAAATATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGA 593

Db       2583 ATTTTTTAAAAATTTTATTTAAATTTTTAAATTTTTTAATTTATTTTATAATTTAATTTA 2524

Qy        594 TCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTTATTT 653

Db       2523 TTTTTTTTTATTTTTCATTAATATTTTAATTATTTTAAAACAAAATAATAAAAAATATAT 2464

Qy        654 AGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGATGGT 713

Db       2463 A-AATAAAAATAAAAAAAAATTAATTTTTAAAAATAAATAAATTATTTTTTTTTTAATTT 2405

Qy        714 TTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCATATACCTAAACT 773

Db       2404 TCAAAAAATAATAAAATTTAATAAATTATTTATAAAAATAAAAAATTATATTTAAAAATT 2345

Qy        774 TGAACAAATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACCTTATTTGAA 831

Db       2344 AAATAAAATTTATTTAAATACAAAAATTAATAAATTATTTTAAAAATAAATAAATTAA 2287


RESULT 8
US-10-473-126-240/c
; Sequence 240, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Methods and nucleic acids for the analysis of  hematopoietic cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/473,126
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 240
;   LENGTH: 8056
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-240

  Query Match             5.1%;  Score 56.6;  DB 8;  Length 8056;
  Best Local Similarity   43.6%;  Pred. No. 0.37;
  Matches  459;  Conservative    0;  Mismatches  579;  Indels   15;  Gaps    4;

Qy          6 TTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAAAGTAT 65

Db       2466 TATAAATAAAAATAAAAAAAAAATTAATTTTTAAAAATAAATAAATTATTTTTTTTTTAAT 2407

Qy         66 ATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAGATTTAATGA 125

Db       2406 TTTCGAAAAATAATAAAATTTAATAAATTATTTATAAAAATAAAAAATTATATTTA---- 2351

Qy        126 GCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTTATTAA 185

Db       2350 --AAAATTAAATAAAATTTATTTAAATACGAAAATTAATAAATTATTTTAAAAATAAATA 2293

Qy        186 AAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCC 245

Db       2292 AATTAAAAAATTAAAATTTTTTTAATTTATTTTAAATATTAAAAAATAAATTTTTATTAA 2233

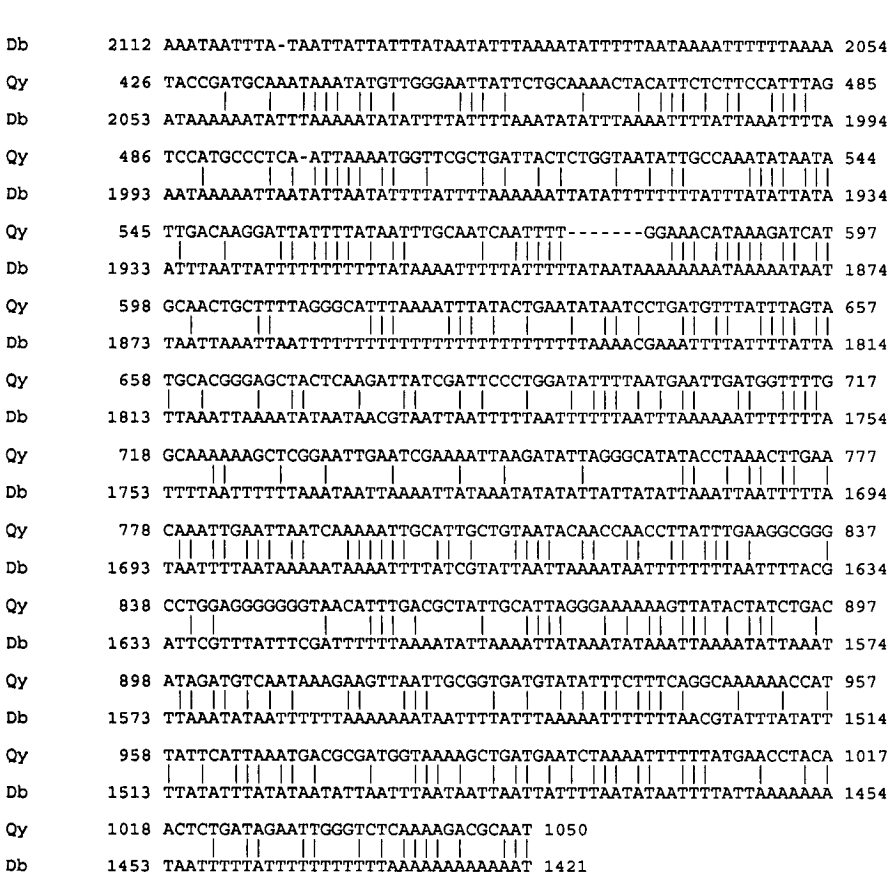
Qy        246 TTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGA 305

Db       2232 AATTAAATTTTTTTTTTTTTTTTATTTTTTTTTTCGTTTTTTTAAATTTTTAATAAAATTA 2173

Qy        306 CTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGT 365

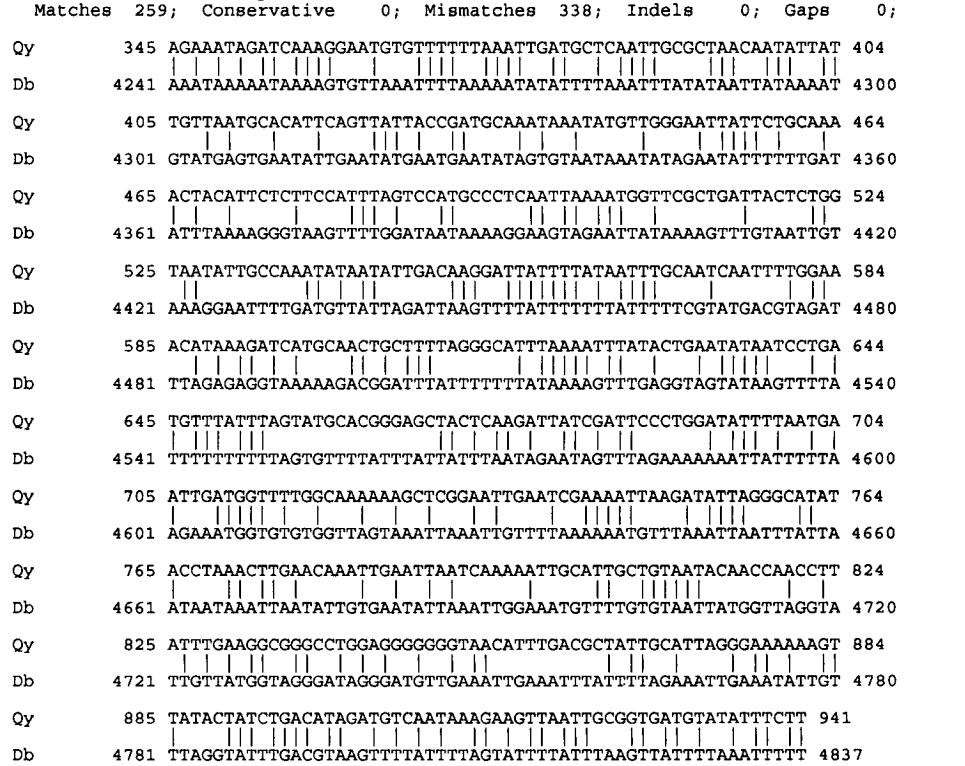
Db       2172 TAATTTTTTTTTTTTTTATTTTTATTTTTAATAAAATAAAAAATTATTAAAATTATATT 2113

Qy        366 GTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTAT 425
```



RESULT 9
US-10-311-455-598
; Sequence 598, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 598
; LENGTH: 16217
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 3949
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-598

Query Match 5.1%; Score 56.2; DB 6; Length 16217;
Best Local Similarity 43.4%; Pred. No. 0.59;
Matches 259; Conservative 0; Mismatches 338; Indels 0; Gaps 0;



RESULT 10
US-09-754-853A-2
; Sequence 2, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 2
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:

```
;   NAME/KEY: CDS
;   LOCATION: (45163)..(45314),(45450)..(45509),(46941)..(48763),(48975)..(49573)
;   OTHER INFORMATION: Clone ID: 240O17_region_G3
US-09-754-853A-2

  Query Match             5.1%;  Score 56.2;  DB 3;  Length 335913;
  Best Local Similarity   46.7%;  Pred. No. 1.9;
  Matches  178;  Conservative    0;  Mismatches  203;  Indels    0;  Gaps    0;

Qy        254 GTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAAC 313
              || ||   | |   ||||| |||| | || ||||   |   |||  || ||    ||  |
Db     223638 GTTTTATGTGGTTCAATTTTAATATAGAATCATGTGTGATTTAGTTTGATTCAATTTGAT 223697

Qy        314 ACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTA 373
                | ||       || | | | | || | | | || || ||   |   | |   | | |||
Db     223698 GATTTTTTATATTTTTAATATATTGTCAACTAAAACTAAATTTATATAGTTATTATGTTA 223757

Qy        374 AATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATG 433
              ||||  |    || ||   | |   | || |  | ||       || |  ||||   | ||
Db     223758 AATTATTTTAAAAATGAATTTATTTTTTTTTTATTAAAATTTGTTTAATATATTTTAGTTG 223817

Qy        434 CAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCC 493
              ||||  | ||| ||      ||||  |||| |  | |    |   |      ||   | |
Db     223818 AAAATTATTATTTTTACTTCTATTTGGCAATAACATAAAATAATATTACTAACTCTTTTT 223877

Qy        494 CTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGG 553
              | |  ||| |    |   ||  |  | || || |         | ||| || |   |
Db     223878 GTAACATAATAAAAATTTGTGTATCATTTGATATTTAATATTTTTTTAAAATAGTATATC 223937

Qy        554 ATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGG 613
              || ||| | |      | | || | |  |   || |  |     |||    |  ||
Db     223938 CTTTTTTAACATAAAAAAGTAAAATATATATTAATTATAAAATAACAAAAATTAATAAAT 223997

Qy        614 CATTTAAAATTTATACTGAAT 634
              ||| || | ||||  | |||
Db     223998 TATTCAATACTTATTATTAAT 224018


RESULT 11
US-09-754-853A-3
; Sequence 3, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
;  APPLICANT: Hauge, Brian M.
;  APPLICANT:  Parnell, Laurence D.
;  APPLICANT:  Parsons, Jeremy D.
;  APPLICANT:  Wang, Ming Li
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Soybean Cyst Nematode Resistance
;  FILE REFERENCE: 38-10(15810)B
;  CURRENT APPLICATION NUMBER: US/09/754,853A
;  CURRENT FILING DATE:  2001-01-05
;  PRIOR APPLICATION NUMBER: US 60/174,880
;  PRIOR FILING DATE: 2000-01-07
;  NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 3
;   LENGTH: 335913
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (46798)..(48763),(48975)..(49573)
;   OTHER INFORMATION: Clone ID: 240O17_region_G3
US-09-754-853A-3

  Query Match             5.1%;  Score 56.2;  DB 3;  Length 335913;
  Best Local Similarity   46.7%;  Pred. No. 1.9;
  Matches  178;  Conservative    0;  Mismatches  203;  Indels    0;  Gaps    0;

Qy        254 GTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAAC 313
              || ||   | |   ||||| |||| | || ||||   |   |||  || ||    ||  |
Db     223638 GTTTTATGTGGTTCAATTTTAATATAGAATCATGTGTGATTTAGTTTGATTCAATTTGAT 223697

Qy        314 ACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTA 373
                | ||       || | | | | || | | | || || ||   |   | |   | | |||
Db     223698 GATTTTTTATATTTTTAATATATTGTCAACTAAAACTAAATTTATATAGTTATTATGTTA 223757

Qy        374 AATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATG 433
              ||||  |    || ||   | |   | || |  | ||       || |  ||||   | ||
Db     223758 AATTATTTTAAAAATGAATTTATTTTTTTTTTATTAAAATTTGTTTAATATATTTTAGTTG 223817

Qy        434 CAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCC 493
              ||||  | ||| ||      ||||  |||| |  | |    |   |      ||   | |
Db     223818 AAAATTATTATTTTTACTTCTATTTGGCAATAACATAAAATAATATTACTAACTCTTTTT 223877

Qy        494 CTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGG 553
              | |  ||| |    |   ||  |  | || || |         | ||| || |   |
Db     223878 GTAACATAATAAAAATTTGTGTATCATTTGATATTTAATATTTTTTTAAAATAGTATATC 223937

Qy        554 ATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGG 613
              || ||| | |      | | || | |  |   || |  |     |||    |  ||
Db     223938 CTTTTTTAACATAAAAAAGTAAAATATATATTAATTATAAAATAACAAAAATTAATAAAT 223997

Qy        614 CATTTAAAATTTATACTGAAT 634
              ||| || | ||||  | |||
Db     223998 TATTCAATACTTATTATTAAT 224018
```

```
Qy        314 ACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTA 373
                | ||       || | | | | || | | | || || ||   |   | |   | | |||
Db     223698 GATTTTTTATATTTTTAATATATTGTCAACTAAAACTAAATTTATATAGTTATTATGTTA 223757

Qy        374 AATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGATG 433
              ||||  |    || ||   | |   | || |  | ||       || |  ||||   | ||
Db     223758 AATTATTTTAAAAATGAATTTATTTTTTTTTTATTAAAATTTGTTTAATATATTTTAGTTG 223817

Qy        434 CAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCC 493
              ||||  | ||| ||      ||||  |||| |  | |    |   |      ||   | |
Db     223818 AAAATTATTATTTTTACTTCTATTTGGCAATAACATAAAATAATATTACTAACTCTTTTT 223877

Qy        494 CTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGG 553
              | |  ||| |    |   ||  |  | || || |         | ||| || |   |
Db     223878 GTAACATAATAAAAATTTGTGTATCATTTGATATTTAATATTTTTTTAAAATAGTATATC 223937

Qy        554 ATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGG 613
              || ||| | |      | | || | |  |   || |  |     |||    |  ||
Db     223938 CTTTTTTAACATAAAAAAGTAAAATATATATTAATTATAAAATAACAAAAATTAATAAAT 223997

Qy        614 CATTTAAAATTTATACTGAAT 634
              ||| || | ||||  | |||
Db     223998 TATTCAATACTTATTATTAAT 224018


RESULT 12
US-10-311-455-516
; Sequence 516, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 516
;   LENGTH: 11790
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-516

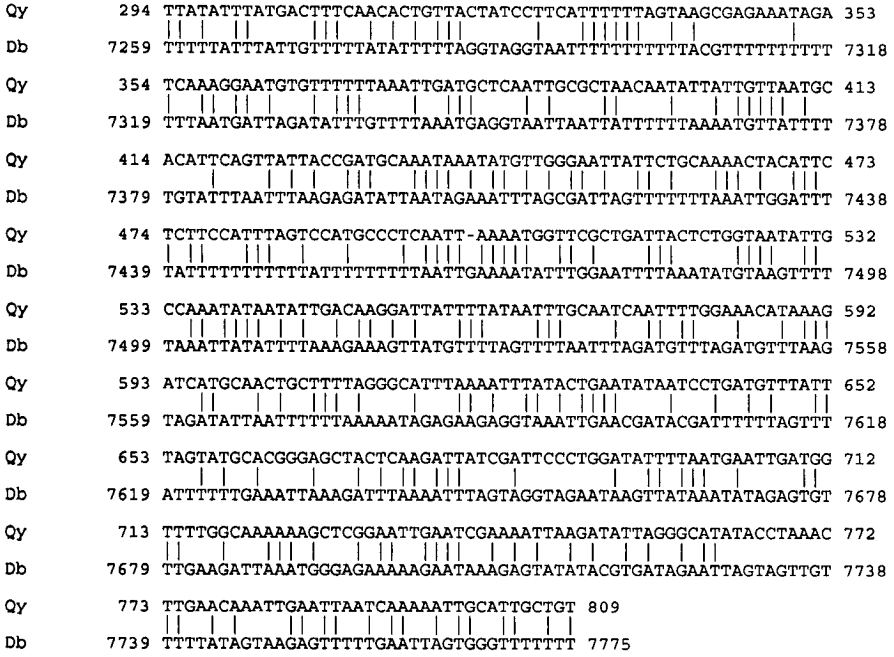
  Query Match             5.0%;  Score 55.4;  DB 6;  Length 11790;
  Best Local Similarity   42.9%;  Pred. No. 0.76;
  Matches  325;  Conservative    0;  Mismatches  431;  Indels    1;  Gaps    1;

Qy         54 TTTTAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAA 113
              |||||| | |   | ||    |  | |  ||||||  ||        |  |||     |
Db       7019 TTTTAATATTTGGTAAATTAAATATATTTAGAGGGTTAGAAAATAGGGATTTTTTAGGAT 7078

Qy        114 TAGATTTAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCT 173
              | | |   |      |||   |   |  |   |||         | | | | | |  | |
Db       7079 TTGTTAATAAATTTAATTAAAGGAAATTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAT 7138

Qy        174 TACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGA 233
              ||  ||||  | |  |||    |        ||| |    ||||   | |   || | |
Db       7139 TATTTTTAAGAGAGATAAATTTTTGAAAAATTAGTTTATTTTTTGAATTTTTGAATATAT 7198

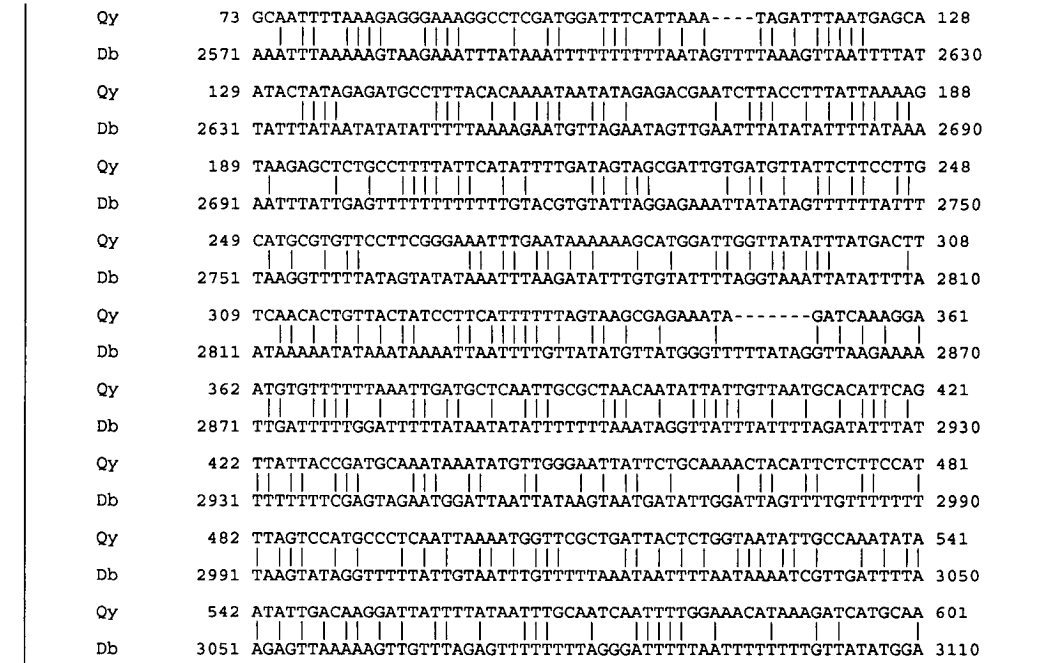
Qy        234 TGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGG 293
              | | || ||    |  ||  || |   |       | |||  ||| | | |   |     |
Db       7199 GGGTTTTTTTTAATTGATTTGTTTAAATATTTATAGTTTTTATAGATAGTTAAAAGATAA 7258
```



RESULT 13
US-10-311-455-1571
; Sequence 1571, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1571
; LENGTH: 5487
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1571

Query Match 4.9%; Score 54.2; DB 6; Length 5487;
Best Local Similarity 45.0%; Pred. No. 1;
Matches 294; Conservative 0; Mismatches 348; Indels 11; Gaps 2;





RESULT 14
US-10-311-455-836
; Sequence 836, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 836
; LENGTH: 6132
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-836

Query Match 4.9%; Score 54; DB 6; Length 6132;
Best Local Similarity 46.2%; Pred. No. 1.1;
Matches 253; Conservative 0; Mismatches 290; Indels 5; Gaps 2;

```
Qy        173 TTACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTG 232
              ||   ||| |  | | | |  | | |       || |    |||  ||| | | |||| |
Db       2476 TTTGTTTTTTATAGATTGATTGTTATTATAAGTAGTGTGTTTTAAATATTTGGGATTATA 2535

Qy        233 ATGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTG 292
               |||   | |    |   |   | |   || |   ||||| |  ||  ||| ||   ||
Db       2536 TTGTATATATAATGTAATTTATTTTATTTTTGATTAATTTTATAAATTAAGTATTTTTTT 2595

Qy        293 GTTATATTTATGACTTTCA---ACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAA 349
               | |   ||  | ||| |   |   |  | ||   ||  |||||||  |   | |||
Db       2596 TATGTTCGTAGAAATTTTAGTTAATATAATTATAGTTTTTTTTTTTTTAGGTATTAAAAAA 2655

Qy        350 TAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTA 409
              |  |  |    | |   ||||  | | |   |  | ||  |      |  | ||  |  |||
Db       2656 TTTAATATGTAAGTAATTTTTAAATAATTTTAGTAAAGCGAAAATAATAAATATTATTTA 2715

Qy        410 ATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTAC 469
              ||       | | |||||    ||      | ||| | |||  |  | || |   | | |
Db       2716 ATTTTTTATTATTTATTGATAATTATTTTTAATGTTTTGAAATATGTTTTAGTATATTTT 2775

Qy        470 ATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATA 529
               || | |   || |     |      |||||||| ||     || | | | |      ||
Db       2776 TTTTTATGTAATATTTGTATTTTTTAAAATTAAAAGGGGATTATGTTAAATATAAGTTTA 2835

Qy        530 TTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATA 589
              |      | |  || | |        |   || |||||| |  ||| |||| |   ||  |
Db       2836 TGATTTATTTAAAAAATTGAGGTAAAATTTATATAATATATAATTAATTATTTTAAAGT- 2894

Qy        590 AAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCTGATGTTT 649
               |  | ||  |      |||||   |||||||||||| || | | |||  |       |||
Db       2895 -ATTTAATTTAGTGATATTTAGTATATTTAAAATTTGTAATTATTATTTTTGGATATTTT 2953

Qy        650 ATTTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAATGAATTGA 709
              || ||    | |    |      |    || | |       ||||||   || | | |
Db       2954 ATATAAACGGAATTATATAATATATGATTTGTTGTGGTTGGTTATATTTATTTGTAATAA 3013

Qy        710 TGGTTTTG 717
              || |||||
Db       3014 TGTTTTTG 3021
```

```
RESULT 15
US-11-097-143-16492/c
; Sequence 16492, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16492
;   LENGTH: 11143
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-16492

  Query Match             4.8%;  Score 53.4;  DB 10;  Length 11143;
  Best Local Similarity   45.9%;  Pred. No. 1.9;
  Matches  218;  Conservative    0;  Mismatches  256;  Indels    1;  Gaps    1;
```

```
Qy        134 ATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACCTTTATTAAAAGTAAGA 193
              | ||| |    | |  | |||| || |   | ||| ||      || ||||||  | |
Db       4747 AGAGAAACATTTATTTAAAAAAAAAAAGTAAAACGGATACATTTTTGATTAAATCTGATT 4688

Qy        194 GCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGC 253
               |||   | ||||  || | | ||  |    ||| | ||  |  |  |  ||
Db       4687 ATTCTTTTTAGTATTGGTACTGTTATTATTATTATTATTATTATTATTATTATTATT 4628

Qy        254 GTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTATATTTATGACTTTCAAC 313
              |  |  ||      |||   || | |   |     |  |||| |||| | | | |
Db       4627 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 4568

Qy        314 ACTGTTACTATCCTT-CATTTTTTAGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTT 372
              | | |   |||  ||  | || ||| ||        ||   |  |  | | | |  | ||
Db       4567 ATTATGGTTATTATTATAATTATTATTATTATTATTATTATTATTATTATTATCATT 4508

Qy        373 AAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACATTCAGTTATTACCGAT 432
              ||  | ||  | |||          |||||||   | ||  |   | |  |        ||
Db       4507 AATATTATTATTAATATAATATTATATATTATATATTATATAATATAATATTAATATTAT 4448

Qy        433 GCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGC 492
              | || | |||  |    ||||||  | |  |     ||| |  |  | | |  |
Db       4447 AAATATTATTATTATTATTATTATTGTTCTTATTATTATTATTATTATTATAATTAT 4388

Qy        493 CCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAG 552
              |    |   | |  |    | |||| | |  | || ||     ||||| ||| | | ||
Db       4387 TATTGTTATTATTAATATTATTATTATTATCATTATTTTTATTAATATTATAATAAGAAT 4328

Qy        553 GATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTT 607
              ||||  | | |   ||    | | | |||  | |||||| |  |  |||  |  ||
Db       4327 AATTAATATTTTTATTATTTTAAGTATTGTTATCATAAATAAAAATCAAGTATTT 4273
```

Search completed: May 11, 2006, 07:08:35
Job time : 741.653 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:23:58 ; Search time 572.043 Seconds
(without alignments)
7892.621 Million cell updates/sec

Title: US-09-423-093B-2_COPY_5365_6471
Perfect score: 1107
Sequence: 1 attcttttaccgagaaatga.....aaattgaatcccgatcttaa 1107

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	49.8	4.5	1134	11	US-10-932-182A-166160	Sequence 166160,
	2	49.8	4.5	1134	11	US-10-932-182A-166160	Sequence 166160,
	3	49.4	4.5	8961	10	US-10-240-708-80	Sequence 80, Appl
c	4	46.6	4.2	584	7	US-09-925-065A-916020	Sequence 916020,
	5	46.6	4.2	161874	17	US-11-121-086-75	Sequence 75, Appl
c	6	46.2	4.2	584	7	US-09-925-065A-894399	Sequence 894399,
c	7	46.2	4.2	584	7	US-09-925-065A-894400	Sequence 894400,
	8	44.6	4.0	544	7	US-09-925-065A-157837	Sequence 157837,
	9	44.6	4.0	547	12	US-10-301-480-250920	Sequence 250920,
	10	44.6	4.0	547	12	US-10-301-480-864329	Sequence 864329,
	11	44.6	4.0	892	7	US-09-925-065A-59025	Sequence 59025, A
	12	44.6	4.0	892	11	US-10-301-480-160263	Sequence 160263,
	13	44.6	4.0	892	12	US-10-301-480-773672	Sequence 773672,
	14	44.6	4.0	6583	10	US-10-240-708-26	Sequence 26, Appl
	15	44.2	4.0	544	7	US-09-925-065A-157838	Sequence 157838,
	16	44.2	4.0	547	12	US-10-301-480-250921	Sequence 250921,
	17	44.2	4.0	547	12	US-10-301-480-864330	Sequence 864330,
	18	44.2	4.0	892	7	US-09-925-065A-59026	Sequence 59026, A
	19	44.2	4.0	892	11	US-10-301-480-160264	Sequence 160264,
	20	44.2	4.0	892	12	US-10-301-480-773673	Sequence 773673,
	21	44.2	4.0	173602	17	US-11-121-086-25	Sequence 25, Appl
	22	44	4.0	2160	7	US-09-925-065A-709387	Sequence 709387,
	23	43.8	4.0	440	7	US-09-925-065A-169350	Sequence 169350,
	24	43.8	4.0	440	7	US-09-925-065A-169351	Sequence 169351,
	25	43.8	4.0	441	12	US-10-301-480-261323	Sequence 261323,
	26	43.8	4.0	441	12	US-10-301-480-261324	Sequence 261324,
	27	43.8	4.0	441	12	US-10-301-480-874732	Sequence 874732,
	28	43.8	4.0	441	12	US-10-301-480-874733	Sequence 874733,
	29	43.8	4.0	493	11	US-10-301-480-48232	Sequence 48232, A
	30	43.8	4.0	493	12	US-10-301-480-661641	Sequence 661641,
	31	43.4	3.9	440	7	US-09-925-065A-169352	Sequence 169352,
	32	43.4	3.9	441	12	US-10-301-480-261325	Sequence 261325,
	33	43.4	3.9	441	12	US-10-301-480-874734	Sequence 874734,
	34	43.4	3.9	620	12	US-10-301-480-327557	Sequence 327557,
	35	43.4	3.9	620	12	US-10-301-480-940966	Sequence 940966,
	36	43.4	3.9	639	7	US-09-925-065A-247449	Sequence 247449,
c	37	43.4	3.9	394468	10	US-10-995-561-13473	Sequence 13473, A
c	38	43	3.9	3069	10	US-10-793-626-4415	Sequence 4415, Ap
c	39	43	3.9	3335	10	US-10-793-626-3366	Sequence 3366, Ap
	40	43	3.9	3456	10	US-10-793-626-4026	Sequence 4026, Ap
c	41	43	3.9	3638	10	US-10-793-626-3640	Sequence 3640, Ap
	42	43	3.9	3737	10	US-10-793-626-4196	Sequence 4196, Ap
	43	43	3.9	4136	10	US-10-793-626-4036	Sequence 4036, Ap
c	44	43	3.9	4346	10	US-10-793-626-3702	Sequence 3702, Ap
	45	43	3.9	5152	10	US-10-240-708-74	Sequence 74, Appl

ALIGNMENTS

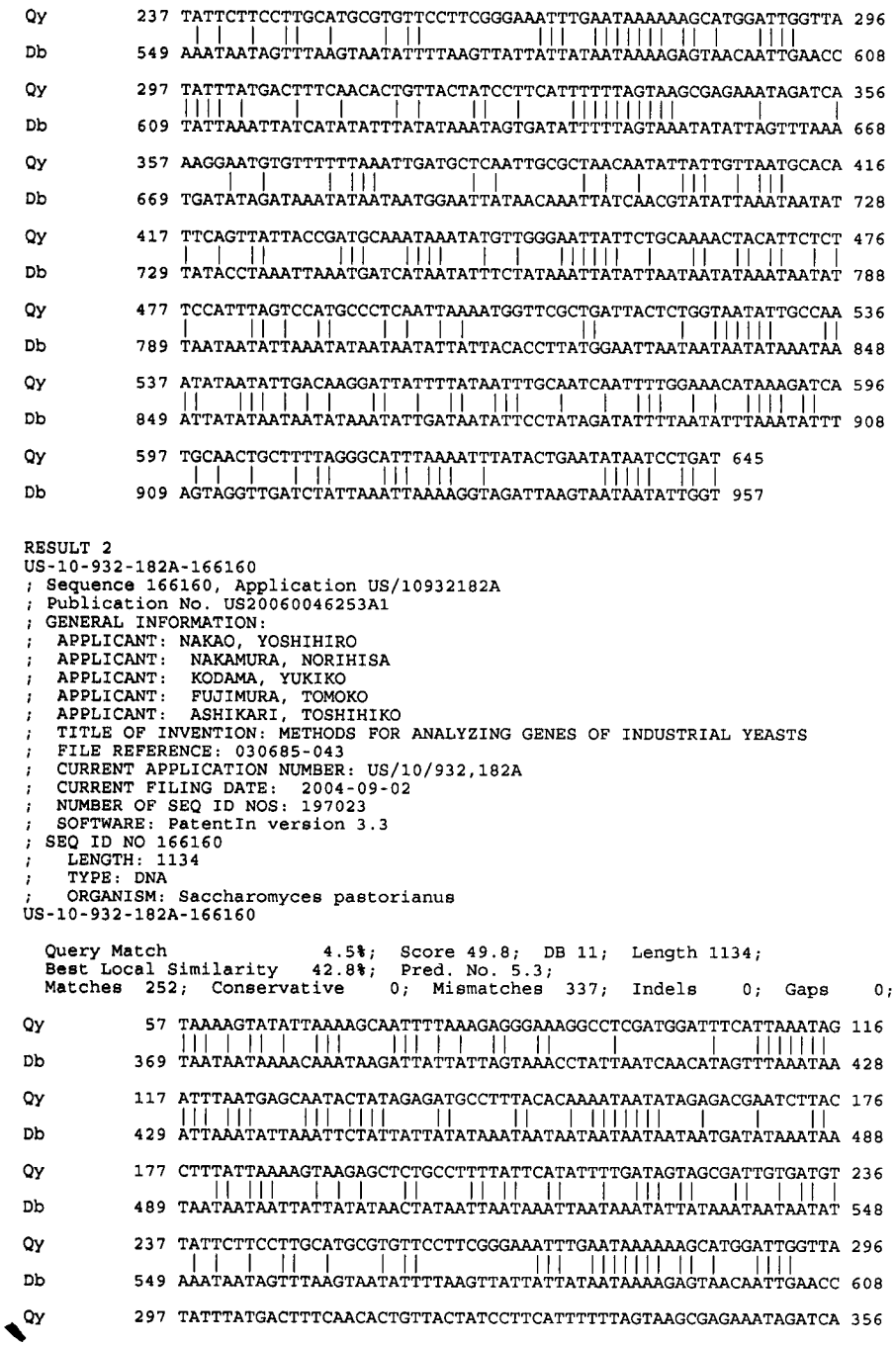
RESULT 1
US-10-932-182A-166160
; Sequence 166160, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166160
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166160

Query Match 4.5%; Score 49.8; DB 11; Length 1134;
Best Local Similarity 42.8%; Pred. No. 5.3;
Matches 252; Conservative 0; Mismatches 337; Indels 0; Gaps 0;

```
Qy        57 TAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAAGGCCTCGATGGATTTCATTAAATAG 116
             ||| | || |  |||    ||| | |  ||  ||     |        |   |||||||
Db       369 TAATAATAAAACAAATAAGATTATTATTAGTAAACCTATTAATCAACATAGTTTAAATAA 428

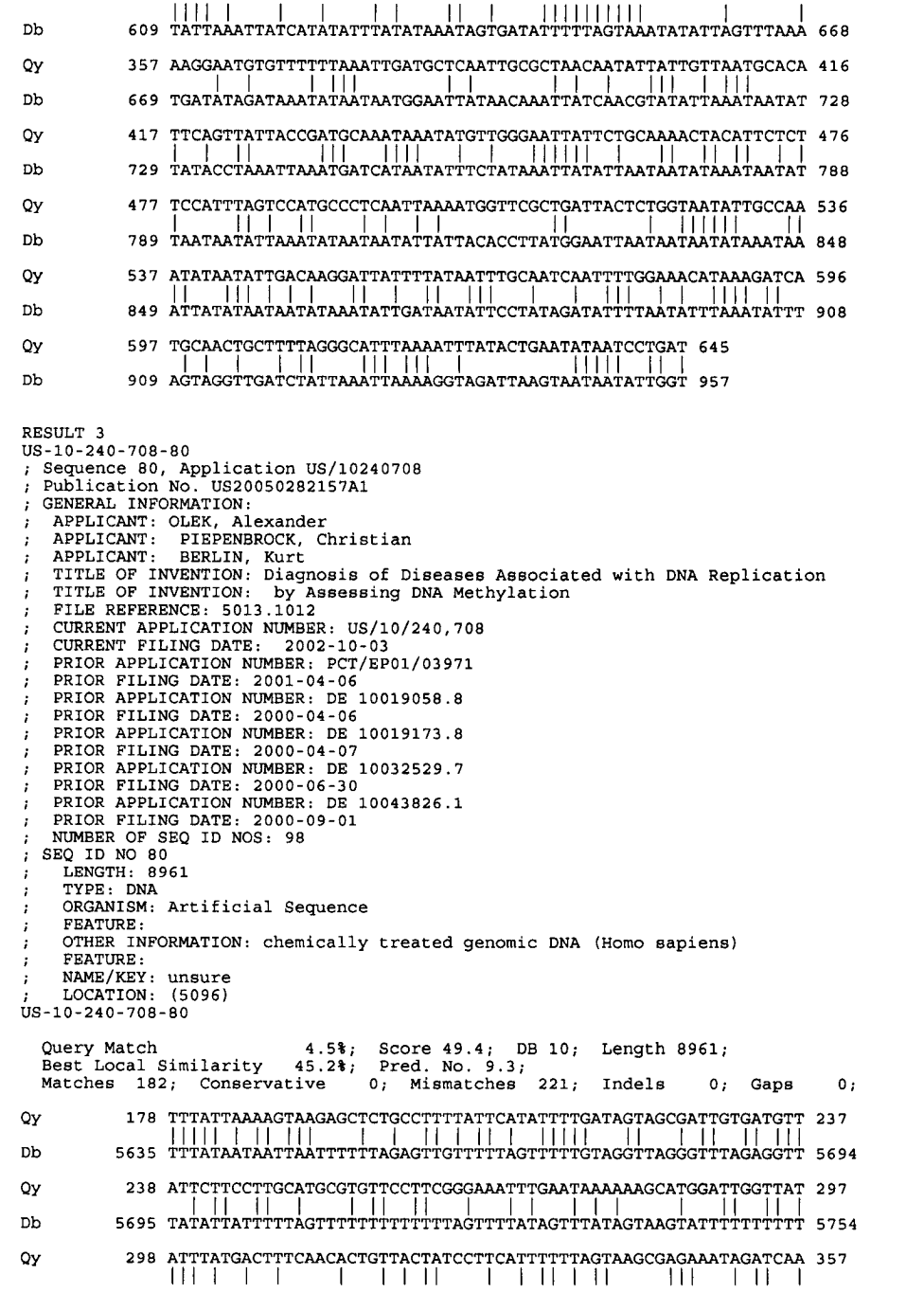
Qy       117 ATTTAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTAC 176
             ||| |||    ||| ||||    ||     ||   | |||||||   |    |    ||
Db       429 ATTAAATATTAAATTCTATTATTATATAAATAATAATAATAATAATAATGATATAAATAA 488

Qy       177 CTTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGT 236
             || |||  | | |  ||   || ||  ||   |  ||| ||   ||  | || |
Db       489 TAATAATAATTATTATATAACTATAATTAATAAATTAATAAATATTATAAATAATAATAT 548
```



RESULT 2
US-10-932-182A-166160
; Sequence 166160, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166160
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166160

Query Match 4.5%; Score 49.8; DB 11; Length 1134;
Best Local Similarity 42.8%; Pred. No. 5.3;
Matches 252; Conservative 0; Mismatches 337; Indels 0; Gaps 0;



RESULT 3
US-10-240-708-80
; Sequence 80, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 80
; LENGTH: 8961
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (5096)
US-10-240-708-80

Query Match 4.5%; Score 49.4; DB 10; Length 8961;
Best Local Similarity 45.2%; Pred. No. 9.3;
Matches 182; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

```
Db       5755 ATTAAAAAGATAATTTATAATGATTAATAGTTTTATTGTCGTTTTATAGATGGATATTTA 5814

Qy        358 AGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACAT 417

Db       5815 ATTTTTTTTTTTTTAATATTGGAAAGTTATTTTATAATAATTTTATTTTTTTATTTATAT 5874

Qy        418 TCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTT 477

Db       5875 TTTGATATTTTATAATAATTGTAAGGAGTTCGTGGTTAATTAAATATATTATAATATTTT 5934

Qy        478 CCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAA 537

Db       5935 TAATATAATTTAGTGTTATATTTTATAATGTTTTATTTTTATTTTATTATTTTTATATTT 5994

Qy        538 TATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTT 580

Db       5995 TTCGAGTTAGGTATTTTATAATATTTATATTTTTTTTATTTTT 6037


RESULT 4
US-09-925-065A-916020/c
; Sequence 916020, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 916020
;   LENGTH: 584
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-916020

  Query Match             4.2%;  Score 46.6;  DB 7;  Length 584;
  Best Local Similarity   51.8%;  Pred. No. 19;
  Matches  132;  Conservative    0;  Mismatches  119;  Indels    4;  Gaps    1;

Qy        583 AAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCT 642

Db        440 AAGTTTAAGGCTGAATCTTGATCTTAAATTGGATTTAAATTATATACGTAATAGATATCT 381

Qy        643 GATGTTTATTTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAAT 702

Db        380 TATTTTTATCTTATTAGCAAGTGATCATGAAAATATTTTTACTTCTAGTGATATTACAT- 322

Qy        703 GAATTGATGGTTTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCAT 762

Db        321 ---TTGACCGTAATATGAAAAAGAAGAGTGATAGCAATACAAATGATGATGTGAGGGCAT 265

Qy        763 ATACCTAAACTTGAACAAATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACC 822

Db        264 ATGCAAATTATCTCCTCCATATAAACAATTAAAAATGTGCAAAAATAATCAGAATCAGCT 205

Qy        823 TTATTTGAAGGCGGG 837

Db        204 CTTTTAAAACTCTGG 190


RESULT 5
US-11-121-086-75
; Sequence 75, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 75
;   LENGTH: 161874
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-75

  Query Match             4.2%;  Score 46.6;  DB 17;  Length 161874;
  Best Local Similarity   44.7%;  Pred. No. 54;
  Matches  181;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy        393 TAACAATATTATTGTTAATGCACATTCAGTTATTACCGATGCAAATAAATATGTTGGGAA 452

Db     125426 TATTTAAATTATATTTTATTTTTAATATTTTATTTTTAATATATTAAAATATATCTTAAA 125485

Qy        453 TTATTCTGCAAAACTACATTCTCTTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGC 512

Db     125486 ATATATTATAAAACATATTTATCTTAAAATATGTTATAAAACGTATTTATCTTAAAATAT 125545

Qy        513 TGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCAA 572

Db     125546 ATTATAAAATGTATTTATCTTAAAATATATTATAAAACGTATTTATCTTAAAATATATTA 125605

Qy        573 TCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGA 632

Db     125606 TAAAACGTATTTATCTTAAAATATATTATAAAATATATTTATTTTAATATATTAAAATCT 125665

Qy        633 ATATAATCCTGATGTTTATTTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGG 692

Db     125666 ATTTTAATATGTTGTTAAAATATATTTCTTTTTAATGTGTTAAAATATATGCTTAAGATA 125725

Qy        693 ATATTTTAATGAATTGATGGTTTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAGAT 752

Db     125726 TATTCATATTTAAATATAAAAGGAGGAAAGAAATTTTAAATTTAATTTTCATTTTAATTT 125785

Qy        753 ATTAGGGCATATACCTAAACTTGAACAAATTGAATTAATCAAAAA 797

Db     125786 TTATTGCATTTTACATATGTATCAGTGACTGAAATTAAAAAAAAA 125830


RESULT 6
US-09-925-065A-894399/c
; Sequence 894399, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
```

```
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 894399
;   LENGTH: 584
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-894399

  Query Match             4.2%;  Score 46.2;  DB 7;  Length 584;
  Best Local Similarity   51.4%;  Pred. No. 22;
  Matches  131;  Conservative    1;  Mismatches  119;  Indels    4;  Gaps    1;

Qy        583 AAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCT 642
              ||   ||| | | |   |     |||  |  | ||||||| | |||||  |||| |   ||
Db        440 AAGTTTAAGGCTGAATCTTGATCTTAAATTGGATTTAAATTATATACGTAATAGATATCT 381

Qy        643 GATGTTTATTTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAAT 702
              || ||||| |  |  ||| | || |     || ||| |   |||    ||||||  |
Db        380 TATTTTTATCTTATTAGCAAGTGATCATGAAAATATTTTTACTTCTAGTGATATTACAT- 322

Qy        703 GAATTGATGGTTTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCAT 762
              ||||  ||  |  |||||     |  || | |    |||| | ||| | |||||||
Db        321 ---TTGACCGTAATATGAAAAAGAAGAGTGATAGCAATACAAATGATGATGTGAGGGCAT 265

Qy        763 ATACCTAAACTTGAACAAATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACC 822
              |: |  |   |       ||  ||  ||| ||||||          |  | || || |
Db        264 AKGCAAATTATCTCCTCCATATAAACAATTAAAAATGTGCAAAAATAATCAGAATCAGCT 205

Qy        823 TTATTTGAAGGCGGG 837
              | ||  ||  | ||
Db        204 CTTTTAAAACTCTGG 190


RESULT 7
US-09-925-065A-894400/c
; Sequence 894400, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 894400
;   LENGTH: 584
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-894400

  Query Match             4.2%;  Score 46.2;  DB 7;  Length 584;
  Best Local Similarity   51.4%;  Pred. No. 22;
  Matches  131;  Conservative    1;  Mismatches  119;  Indels    4;  Gaps    1;

Qy        583 AAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTATACTGAATATAATCCT 642
              ||   ||| | | |   |     |||  |  | ||||||| | |||||  |||| |   ||
Db        440 AAGTTTAAGGCTGAATCTTGATCTTAAATTGGATTTAAATTATATACGTAATAGATATCT 381

Qy        643 GATGTTTATTTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATATTTTAAT 702
              || ||||| |  |  ||| | || |     || ||| |   |||    ||||||  |
Db        380 TATTTTTATCTTATTAGCAAGTGATCATGAAAATATTTTTACTTCTAGTGATATTACAT- 322

Qy        703 GAATTGATGGTTTTGGCAAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTAGGGCAT 762
              ||||  ||  |  |||||     |  || | |    |||| | ||| | |||||||
Db        321 ---TTGACCGTAATATGAAAAAGAAGAGTGATAGCAATACAAATGATGATGTGAGGGCAT 265

Qy        763 ATACCTAAACTTGAACAAATTGAATTAATCAAAAATTGCATTGCTGTAATACAACCAACC 822
              || |  |   |       ||  ||  ||| ||||||          |  | || || |
Db        264 ATGCAAATTATCTCCTCCATATAAACAATTAAAAATGTGCAAAAATAATCAGAATCAGCT 205

Qy        823 TTATTTGAAGGCGGG 837
              | ||  ||  | ||
Db        204 CTTTTAAAACTCTGG 190


RESULT 8
US-09-925-065A-157837
; Sequence 157837, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 157837
;   LENGTH: 544
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-157837

  Query Match             4.0%;  Score 44.6;  DB 7;  Length 544;
  Best Local Similarity   60.2%;  Pred. No. 43;
  Matches   74;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        499 TTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTAT 558
              || ||| |||    | || |||    || || | | ||||||||    | ||| |||| |
Db        340 TTTAAAAGGTAAAATCATGACTTATATATTAGTACAAAATATAAGTGAGTCAAAGATTTT 399

Qy        559 TTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTT 618
              || || |  | ||   ||   |||| ||  ||||| ||  ||||| ||  || | |
Db        400 ATTTTACTGAGTAACTGATAAGGGAACCAGCAAGATGATAAAACTGATTCAAGAGATTCC 459

Qy        619 AAA 621
              |||
Db        460 AAA 462


RESULT 9
US-10-301-480-250920
; Sequence 250920, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
```

```
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 250920
;   LENGTH: 547
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-250920

  Query Match             4.0%;  Score 44.6;  DB 12;  Length 547;
  Best Local Similarity   60.2%;  Pred. No. 43;
  Matches   74;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        499 TTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTAT 558
              || ||| |||    | || |||    || || | | ||||||||    | ||| |||| |
Db        340 TTTAAAAGGTAAAATCATGACTTATATATTAGTACAAAATATAAGTGAGTCAAAGATTTT 399

Qy        559 TTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTT 618
               || || |  | ||   ||   |||| ||  ||||| ||  ||||| ||  || |  |
Db        400 ATTTTACTGAGTAACTGATAAGGGAACCAGCAAGATGATAAAACTGATTCAAGAGATTCC 459

Qy        619 AAA 621
              |||
Db        460 AAA 462


RESULT 10
US-10-301-480-864329
; Sequence 864329, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 864329
;   LENGTH: 547
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-864329

  Query Match             4.0%;  Score 44.6;  DB 12;  Length 547;
  Best Local Similarity   60.2%;  Pred. No. 43;
  Matches   74;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        499 TTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTAT 558
              || ||| |||    | || |||    || || | | ||||||||    | ||| |||| |
Db        340 TTTAAAAGGTAAAATCATGACTTATATATTAGTACAAAATATAAGTGAGTCAAAGATTTT 399

Qy        559 TTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTT 618
               || || |  | ||   ||   |||| ||  ||||| ||  ||||| ||  || |  |
Db        400 ATTTTACTGAGTAACTGATAAGGGAACCAGCAAGATGATAAAACTGATTCAAGAGATTCC 459

Qy        619 AAA 621
              |||
Db        460 AAA 462


RESULT 11
US-09-925-065A-59025
; Sequence 59025, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 59025
;   LENGTH: 892
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-59025

  Query Match             4.0%;  Score 44.6;  DB 7;  Length 892;
  Best Local Similarity   60.2%;  Pred. No. 48;
  Matches   74;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        499 TTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTAT 558
              || ||| |||    | || |||    || || | | ||||||||    | ||| |||| |
Db        763 TTTAAAAGGTAAAATCATGACTTATATATTAGTACAAAATATAAGTGAGTCAAAGATTTT 822

Qy        559 TTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTT 618
               || || |  | ||   ||   |||| ||  ||||| ||  ||||| ||  || |  |
Db        823 ATTTTACTGAGTAACTGATAAGGGAACCAGCAAGATGATAAAACTGATTCAAGAGATTCC 882

Qy        619 AAA 621
              |||
Db        883 AAA 885


RESULT 12
US-10-301-480-160263
; Sequence 160263, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 160263
;   LENGTH: 892
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-160263

  Query Match             4.0%;  Score 44.6;  DB 11;  Length 892;
  Best Local Similarity   60.2%;  Pred. No. 48;
  Matches   74;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        499 TTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTAT 558
              || ||| |||    | || |||    || || | | ||||||||    | ||| |||| |
Db        763 TTTAAAAGGTAAAATCATGACTTATATATTAGTACAAAATATAAGTGAGTCAAAGATTTT 822
```

```
Qy        559 TTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTT 618
              || || |  | ||   ||   |||| ||  ||||| ||  ||||| ||  || |  |
Db        823 ATTTTACTGAGTAACTGATAAGGGAACCAGCAAGATGATAAAACTGATTCAAGAGATTCC 882

Qy        619 AAA 621
              |||
Db        883 AAA 885
```

RESULT 13
US-10-301-480-773672
; Sequence 773672, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 773672
; LENGTH: 892
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-773672

Query Match 4.0%; Score 44.6; DB 12; Length 892;
Best Local Similarity 60.2%; Pred. No. 48;
Matches 74; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```
Qy        499 TTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTAT 558
              || ||| |||    | || |||    || || | | ||||||||    | ||| |||| |
Db        763 TTTAAAAGGTAAAATCATGACTTATATATTAGTACAAAATATAAGTGAGTCAAAGATTTT 822

Qy        559 TTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTT 618
              || ||| |  | ||   ||   |||| ||  ||||| ||  ||||| ||  || |  |
Db        823 ATTTTACTGAGTAACTGATAAGGGAACCAGCAAGATGATAAAACTGATTCAAGAGATTCC 882

Qy        619 AAA 621
              |||
Db        883 AAA 885
```

RESULT 14
US-10-240-708-26
; Sequence 26, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 26
; LENGTH: 6583
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-26

Query Match 4.0%; Score 44.6; DB 10; Length 6583;
Best Local Similarity 44.2%; Pred. No. 69;
Matches 277; Conservative 0; Mismatches 344; Indels 6; Gaps 2;

```
Qy          2 TTCTTTTACCGAGAAATGATATACATTCTCTTATAAGAGAAAAAGCATTTCCTTTTAAAA 61
              || |||||       |   ||||  ||      ||||      | |   || |  | |  |
Db       5000 TTTTTTTATTATATATATATATGTATATATATATATATATATATATATATTGAT 5059

Qy         62 GTATATTAAAAGCAATTTTAAAGAGGGAAA----GGCCTCGATGGATTTCATTAAATAGA 117
              ||||||    |   | ||||||  || || |     |    | |   ||| | | | |||
Db       5060 GTATATATTAGTAATTTTTAAGAAGAGATATAGGAGTGGGGGTTTTTTTTAGTTAGTAGT 5119

Qy        118 TTTAATGAGCAATACTATAGAGATGCCTTTACACAAAATAATATAGAGACGAATCTTACC 177
               |||   ||   |  |  |     |   |||     ||||  | | | |   ||||
Db       5120 ATTAGGAAGTTTTTTTTAAAGTTTAATTTTTTTTTAAATTAGAGATATGGGAATAAGGGA 5179

Qy        178 TTTATTAAAAGTAAGAGCTCTGCCTTTTATTCATATTTTGATAGTAGCGATTGTGATGTT 237
              |   || ||     | |          | |  |  | || |  ||   | ||||  | | ||
Db       5180 TAATTTTAAGTGTTGGGTAAGTTAGTATCGTTTTGTTATTTTATATGGGATTAGGTTTTT 5239

Qy        238 ATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAATAAAAAAGCATGGATTGGTTAT 297
              ||| |   ||       |  ||| |  ||  ||||| || ||| ||    | |
Db       5240 ATTTTGTTTTTTTGTTAAAATGAATGTGGATAGGTGTTTAAAATAGTATGTATGTAGTGT 5299

Qy        298 ATTTATGACTTTCAACACTGTTACTATCCTTCATTTTTTAGTAAGCGAGAAATAGATCAA 357
              |   |||   | ||     |  |||  |   ||  | | ||     | | |||    |
Db       5300 AAAGGTGAAGATGAAGGAATTGTGTATAGT--GTTGGTGATTATTATATATATATGGAAT 5357

Qy        358 AGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAACAATATTATTGTTAATGCACAT 417
              || |   | ||||||    |   || | | |  || |      | || |  ||||   | |
Db       5358 AGAATATTTTTTTTTTTTTGATTGATTATTAACGGTTTTGGTTTTTTGTTTAAGTAATTT 5417

Qy        418 TCAGTTATTACCGATGCAAATAAATATGTTGGGAATTATTCTGCAAAACTACATTCTCTT 477
              |||                | | |  | || | ||| ||    |  | |  ||
Db       5418 ATTTTTAGGCGTTTATTTTTGTAGTTTAATTTGATTGATTTTGATTGTATTTAATTATTT 5477

Qy        478 CCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAA 537
              || |  |  |     |  | |  |||| ||   || ||| | |  |  ||  |  |
Db       5478 TTATATTTTTAAATTATTTTAATCGAATGATTTTGTGTTTATTGTATTTTTACGGTTATT 5537

Qy        538 TATAATATTGACAAGGATTATTTTATAATTTGCAATCAATTTTGGAAACATAAAGATCAT 597
              | || ||| |  |  |||     | |||| || |||   |||| ||  | |   |  |
Db       5538 TTTATTATGGTAAGTGATATAGTGTTAATATGAAATTTGTTTTTTAAGTAAATTTAAGTT 5597

Qy        598 GCAACTGCTTTTAGGGCATTTAAAATT 624
              | || |   || ||   ||   | | |
Db       5598 GAAAATCGATTAAGTAAATAATAGAAT 5624
```

RESULT 15
US-09-925-065A-157838
; Sequence 157838, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092

```
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 157838
;   LENGTH: 544
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-157838

  Query Match             4.0%;  Score 44.2;  DB 7;  Length 544;
  Best Local Similarity   59.3%;  Pred. No. 52;
  Matches   73;  Conservative    1;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        499 TTAAAATGGTTCGCTGATTACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTAT 558
              || ||| |||    | || |||    || || | | |||||||:    | ||| |||| |
Db        340 TTTAAAAGGTAAAATCATGACTTATATATTAGTACAAAATATAMGTGAGTCAAAGATTTT 399

Qy        559 TTTATAATTTGCAATCAATTTTGGAAACATAAAGATCATGCAACTGCTTTTAGGGCATTT 618
               || || |  | ||   ||   |||| ||  ||||| ||  ||||| ||  || |  |
Db        400 ATTTTACTGAGTAACTGATAAGGGAACCAGCAAGATGATAAAACTGATTCAAGAGATTCC 459

Qy        619 AAA 621
              |||
Db        460 AAA 462


Search completed: May 11, 2006, 08:30:20
Job time : 573.043 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:45:22 ; Search time 2417.44 Seconds
(without alignments)
15660.244 Million cell updates/sec

Title: US-09-423-093B-2_COPY_13156_13821
Perfect score: 666
Sequence: 1 atgaatttgtatggtatttt.....gatcgccaacatctatttaa 666

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	666	100.0	14002	1	AF061251	Escherici
	2	666	100.0	14024	6	BD085555	Nucleic a
	3	666	100.0	14029	6	AX029262	Sequence
c	4	666	100.0	14181	6	BD184789	Nucleic a
c	5	666	100.0	14187	6	AR204225	Sequence
c	6	666	100.0	14187	6	AR637574	Sequence
	7	666	100.0	31482	1	AB008676	Escherich
c	8	666	100.0	110000	1	AE005174_28	Continuation (29 o
c	9	666	100.0	110000	1	BA000007_27	Continuation (28 o
	10	218.2	32.8	687	6	AR318202	Sequence
	11	183.6	27.6	26953	1	ALW243431	AJ243431 Acinetoba
	12	170.8	25.6	110000	1	CR543861_00	CR543861 Acinetoba
c	13	120.6	18.1	205050	1	AL646082	Ralstonia
	14	85.2	12.8	263335	1	CJ11168X5	AL139078 Campyloba
	15	65.8	9.9	10722	1	AE013035	AE013035 Thermoana
	16	61.2	9.2	110000	1	CR626927_08	Continuation (9 of
c	17	59.4	8.9	110000	1	BA000028_29	Continuation (30 o
	18	56.6	8.5	11485	1	AE008469	AE008469 Streptoco
	19	53.6	8.0	10987	1	AE011363	AE011363 Leptospir
c	20	53.6	8.0	302171	1	AE017295	AE017295 Leptospir
	21	52.8	7.9	179553	14	AC024253	AC024253 Homo sapi
	22	52.6	7.9	349723	1	BX842650	BX842650 Bdellovib
c	23	52.4	7.9	110000	1	CR522870_00	CR522870 Desulfota
	24	51.8	7.8	705	1	AY531523	AY531523 Campyloba
	25	51.8	7.8	735	1	AY531522	AY531522 Campyloba
	26	51.8	7.8	110000	1	BA000016_05	Continuation (6 of
	27	51.8	7.8	192187	2	AC117072	AC117072 Dictyoste
	28	51.8	7.8	314150	1	CJ11168X1	AL139074 Campyloba
	29	51.4	7.7	302399	1	AE017164	AE017164 Prochloro
	30	50.4	7.6	6477	1	AF001498	AF001498 Campyloba
c	31	50.4	7.6	110000	1	CP000025_11	Continuation (12 o
	32	50.2	7.5	746	1	AY531511	AY531511 Campyloba
	33	50.2	7.5	746	1	AY531512	AY531512 Campyloba
	34	50.2	7.5	746	1	AY531513	AY531513 Campyloba
	35	50.2	7.5	746	1	AY531514	AY531514 Campyloba
	36	50.2	7.5	746	1	AY531515	AY531515 Campyloba
	37	50.2	7.5	746	1	AY531516	AY531516 Campyloba
	38	50.2	7.5	746	1	AY531517	AY531517 Campyloba
	39	50.2	7.5	746	1	AY531518	AY531518 Campyloba
	40	50.2	7.5	746	1	AY531519	AY531519 Campyloba
	41	50.2	7.5	746	1	AY531520	AY531520 Campyloba
	42	50.2	7.5	746	1	AY531521	AY531521 Campyloba
	43	50.2	7.5	110000	1	CP000025_02	Continuation (3 of
c	44	49	7.4	341050	2	PFA929357	AL929357 Plasmodiu
	45	48.8	7.3	1510	15	AK066657	AK066657 Oryza sat

ALIGNMENTS

RESULT 1
AF061251
LOCUS AF061251 14002 bp DNA linear BCT 20-AUG-1998
DEFINITION Escherichia coli serotype O157:H7 O antigen gene cluster.
ACCESSION AF061251
VERSION AF061251.1 GI:3435170
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 14002)
AUTHORS Wang,L. and Reeves,P.R.
TITLE Organization of Escherichia coli O157 O antigen gene cluster and identification of its specific genes
JOURNAL Infect. Immun. 66 (8), 3545-3551 (1998)
PUBMED 9673232
REFERENCE 2 (bases 1 to 14002)
AUTHORS Wang,L.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1998) Microbiology, The University of Sydney, NSW 2006, Australia
FEATURES Location/Qualifiers
source 1. .14002
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="C664-1992"
/serotype="O157:H7"
/db_xref="taxon:562"
gene 60. .842
/gene="wbdN"
CDS 60. .842
/gene="wbdN"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdN"
/protein_id="AAC32339.1"
/db_xref="GI:3435171"
/translation="MNKETVSIIMPVYNGAKTIISSVESIIHQSYQDFVLYIIDDCST
DDTFSLINSRYKNNQKIRILRNKTNLGVAESRNYGIEMATGKYISFCDADDLWHEKKL
ERQIEVLNNECVDVVCSNYYVIDNNRNIVGEVNAPHVINYRKMLMKNYIGNLTGIYNA

NKLGKFYQKKIGHEDYLMWLEIINKTNGAICIQDNLAYYMRSNNSLSGNKIKAAKWTW
SIYREHLHLSFPKTLYYFLLYASNGVMKKITHSLLRRKETKK"
gene 839. .2023
/gene="wzy"
CDS 839. .2023
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O antigen polymerase Wzy"
/protein_id="AAC32340.1"
/db_xref="GI:3435172"
/translation="MKSAAKLIFLFLFTLYSLQLYGVIIDDRITNFDTKVLTSIIIIF
QIFFVLLFYLTIINERKQQKKFIVNWELKLILVFLFVTIEIAAVVLFLKEGIPIFDDD
PGGAKLRIAEGNGLYIRYIKYFGNIVVFALIILYDEHKFKQRTIIFVYFTTIALFGYR
SELVLLILQYILITNILSKDNRNPKIKRIIGYFLLVGVVCSLFYLSLGQDGEQNDSYN
NMLRIINRLTIEQVEGVPYVVSESIKNDFFPTPELEKELKAIINRIQGIKHQDLFYGE
RLHKQVFGDMGANFLSVTTYGAELLVFFGFLCVFIIPLGIYIPFYLLKRMKKTHSSIN
CAFYSYIIMILLQYLVAGNASAFFFGPFLSVLIMCTPLILLHDTLKRLSRNENISYNC
DL"
gene 1992. .2738
/gene="wbdO"
CDS 1992. .2738
/gene="wbdO"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdO"
/protein_id="AAC32341.1"
/db_xref="GI:3435173"
/translation="MKISVITVTYNNAEGLEKTLSSLSILKIKPFEIIIVDGGSTDGT
NRVISRFTSMNITHVYEKDEGIYDAMNKGRMLAKGDLIHYLNAGDSVIGDIYKNIKEP
CLIKVGLFENDKLLGFSSITHSNTGYCHQGVIFPKNHSEYDLRYKICADYKLIQEVFP
EGLRSLSLITSGYVKYDMGGVSSKKRILRDKELAKIMFEKNKKNLIKFIPISIIKILF
PERLRRVLRKMQYICLTLFFMKNSSPYDNE"
gene 2725. .4116
/gene="wzx"
CDS 2725. .4116
/gene="wzx"
/codon_start=1
/transl_table=11
/product="O antigen flippase Wzx"
/protein_id="AAC32342.1"
/db_xref="GI:3435174"
/translation="MIMNKIKKILKFCTLKKYDTSSALGREQERYRIISLSVISSLIS
KILSLLSLILTVSLTLPYLGQERFGVWMTITSLGAALTFLDLGIGNALTNRIAHSFAC
GKNLKMSRQISGGLTLLAGLSFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKV
FVIIFGLGIYSNGVQKVYMGIQKAYISNIVNAIFILLSIITLVISSKLHAGLPVLIVS
TLGIQYISGIYLTINLIIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSG
DNFIISITLGVTYVAVFSITQRLFQISTVPLTIYNIPLWAAYADAHARNDTQFIKKTL
RTSLKIVGISSFLLAFILVVFGSEVVNIWTEGKIQVPRTFIIAYALWSVIDAFSNTFA
SFLNGLNIVKQQMLAVVTLILIAIPAKYIIVSHFGLTVMLYCFIFIYIVNYFIWYKCS
FKKHIDRQLNIRG"
gene 4113. .5213
/gene="per"
CDS 4113. .5213
/gene="per"
/codon_start=1
/transl_table=11
/product="perosamine synthetase Per"
/protein_id="AAC32343.1"
/db_xref="GI:3435175"
/translation="MKMKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKF
AEQNHVQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFV
DSDNETWQMSVSDIEQKITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEA
FGSKYKGKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHR
QYWHDVIGYNYRMTNICAAIGLAQLEQADDFISRKREIADIYKKNINSLVQVHKESKD
VFHTYWMVSILTRTAEEREELRNHLADKLIETRPVFYPVHTMPMYSEKYQKHPIAEDL
GWRGINLPSFPSLSNEQVIYICESINEFYSDK"
gene 5238. .6452
/gene="wbdP"
CDS 5238. .6452
/gene="wbdP"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdP"
/protein_id="AAC32344.1"
/db_xref="GI:3435176"
/translation="MKIALNSDGFYEWGGGIDFIKYILSILETKPEICIDILLPRNDI
HSLIREKAFPFKSILKAILKRERPRWISLNRFNEQYYRDAFTQNNIETNLTFIKSKSS
AFYSYFDSSDCDVILPCMRVPSGNLNKKAWIGYIYDFQHCYYPSFFSKREIDQRNVFF
KLMLNCANNIIVNAHSVITDANKYVGNYSAKLHSLPFSPCPQLKWFADYSGNIAKYNI
DKDYFIICNQFWKHKDHATAFRAFKIYTEYNPDVYLVCTGATQDYRFPGYFNELMVLA
KKLGIESKIKILGHIPKLEQIELIKNCIAVIQPTLFEGGPGGGVTFDAIALGKKVILS
DIDVNKEVNCGDVYFFQAKNHYSLNDAMVKADESKIFYEPTTLIELGLKRRNACADFL
LDVVKQEIESRS"
gene 6472. .7590
/gene="gmd"
CDS 6472. .7590
/gene="gmd"
/codon_start=1
/transl_table=11
/product="GDP-D-mannose dehydratase Gmd"
/protein_id="AAC32345.1"
/db_xref="GI:3435177"
/translation="MTKVALITGVTGQDGSYLAEFLLDKGYEVHGIKRRASSFNTERI
DHIYQDPHGSNPNFPHLHYGDLTDSSNLTRILKEVQPDEVYNLAAMSHVAVSFESPEYT
ADVDAIGTLRLLEAIRFLGLENKTRFYQASTSELYGLVQEIPQKESTPFYPRSPYAVA
KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRGLANIAQGLESCLYL
GNMDSLRDWGHAKDYVRMQWLMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLGIKMSF
VGKGIEEKGIVDSVEGQDAPGVKPGDVIVAVDPRYFRPAEVDTLLGDPSKANLKLGWR
PEITLAEMISEMVAKDLEAAKKHSLLKSHGFSVSLALE"
gene 7587. .8558
/gene="fcl"
CDS 7587. .8558
/gene="fcl"
/codon_start=1
/transl_table=11
/product="fucose synthetase Fcl"
/protein_id="AAC32346.1"
/db_xref="GI:3435178"
/translation="MMMNKQRIFIAGHQGMVGSAITRRLKQRDDVELVLRTRDELNLL
DSSAVLDFFSSQKIDQVYLAAAKVGGILANSSYPADFIYENIMIEANVIHAAHKNNVN
KLLFLGSSCIYPKLAHQPIMEDELLQGKLEPTNEPYAIAKIAGIKLCESYNRQFGRDY
RSVMPTNLYGPNDNFHPSNSHVIPALLRRFHDAVENNSPNVVVWGSGTPKREFLHVDD
MASASIYVMEMPYDIWQKNTKVMLSHINIGTGIDCTICELAETIAKVVGYKGHITFDT
TKPDGAPRKLLDVTLLHQLGWNHKITLHKGLENTYNWFLENQLQYRG"
gene 8561. .9070
/gene="wbdQ"
CDS 8561. .9070
/gene="wbdQ"
/codon_start=1
/transl_table=11
/product="WbdQ"
/protein_id="AAC32347.1"
/db_xref="GI:3435179"
/translation="MFLHSQDFATIVRSTPLISIDLIVENEFGEILLGKRINRPAQGY
WFVPGGRVLKDEKLQTAFERLTEIELGIRLPLSVGKFYGIWQHFYEDNSMGGDFSTHY
IVIAFLLKLQPNILKLPKSQHNAYCWLSRAKLINDDDVHYNCRAYFNNKTNDAIGLDN
KDIICLMRQ"
gene 9052. .10500
/gene="manC"
CDS 9052. .10500
/gene="manC"
/codon_start=1
/transl_table=11
/product="mannose-1-P guanosyltransferase ManC"
/protein_id="AAC32348.1"
/db_xref="GI:3435180"
/translation="MSDAPIIAVVMAGGTGSRLWPLSRELYPKQFLQLSGDNTLLQTT
LLRLSGLSCQKPLVITNEQHRFVVAEQLREINKLNGNIILEPCGRNTAPAIAISAFHA
LKRNPQEDPLLLVLAADHVIAKESVFCDAIKNATPIANQGKIVTFGIIPEYAETGYGY
IERGELSVPLQGHENTGFYYVNKFVEKPNRETAELYMTSGNHYWNSGIFMFKASVYLE
ELRKFRPDIYNVCEQVASSSYIDLDFIRLSKEQFQDCPAESIDFAVMEKTEKCVVCPV
DIGWSDVGSWQSLWDISLKSKTGDVCKGDILTYDTKNNYIYSESALVAAIGIEDMVIV
QTKDAVLVSKKSDVQHVKKIVEMLKLQQRTEYISHREVFRPWGKFDSIDQGERYKVKK
IIVKPGEGLSLRMHHHRSEHWIVLSGTAKVTLGDKTKLVTANESIYIPLGAAYSLENP
GIIPLNLIEVSSGDYLGEDDIIRQKERYKHED"

```
     gene            10504. .11874
                     /gene="manB"
     CDS             10504. .11874
                     /gene="manB"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphomannomutase ManB"
                     /protein_id="AAC32349.1"
                     /db_xref="GI:3435181"
                     /translation="MKSLTCFKAYDIRGKLGEELNEDIAWRIGRAYGEFLKPKTIVLG
                     GDVRLTSEALKLALAKGLQDAGVDVLDIGMSGTEEIYFATFHLGVDGGIEVTASHNPM
                     DYNGMKLVREGARPISGDTGLRDVQRLAEANDFPPVDETKRGRYQQINLRDAYVDHLF
                     GYINVKNLTPLKLVINSGNGAAGPVVDAIEARFKALGAPVELIKVHNTPDGNFPNGIP
                     NPLLPECRDDTRNAVIKHGADMGIAFDGDFDRCFLFDEKGQFIEGYYIVGLLAEAFLE
                     KNPGAKIIHDPRLSWNTVDVVTAAGGTPVMSKTGHAFIKERMRKEDAIYGGEMSAHHY

  Query Match             100.0%;  Score 666;  DB 1;  Length 14002;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-120;
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13137 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 13196

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13197 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 13256

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13257 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 13316

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13317 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 13376

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13377 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 13436

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13437 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 13496

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13497 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 13556

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13557 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 13616

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13617 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 13676

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13677 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 13736

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13737 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 13796

Qy        661 ATTTAA 666
              ||||||
Db      13797 ATTTAA 13802


RESULT 2
BD085555
LOCUS       BD085555               14024 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085555
VERSION     BD085555.1  GI:22631165
KEYWORDS    JP 2001522248-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 14024)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 2 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Escherichia coli
            PN   JP 2001522248-A/2
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14024
            FT                   /organism='Escherichia coli'.
FEATURES             Location/Qualifiers
     source          1. .14024
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             100.0%;  Score 666;  DB 6;  Length 14024;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-120;
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13156 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 13215

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13216 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 13275

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13276 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 13335

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13336 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 13395

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13396 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 13455

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13456 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 13515

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13516 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 13575

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13576 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 13635

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13636 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 13695

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      13696 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 13755

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13756 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 13815

Qy        661 ATTTAA 666
              ||||||
Db      13816 ATTTAA 13821


RESULT 3
AX029262
LOCUS       AX029262               14029 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 2 from Patent WO9850531.
ACCESSION   AX029262
VERSION     AX029262.1  GI:10190113
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 2 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .14029
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
ORIGIN

  Query Match             100.0%;  Score 666;  DB 6;  Length 14029;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-120;
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13156 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 13215

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13216 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 13275

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13276 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 13335

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13336 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 13395

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13396 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 13455

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13456 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 13515

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13516 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 13575

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13576 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 13635

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13636 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 13695

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13696 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 13755

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13756 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 13815

Qy        661 ATTTAA 666
              ||||||
Db      13816 ATTTAA 13821


RESULT 4
BD184789/c
LOCUS       BD184789               14181 bp    DNA     linear   PAT 17-JUN-2003
DEFINITION  Nucleic acid molecule and polypeptide specific to intestinal
            hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
            thereof.
ACCESSION   BD184789
VERSION     BD184789.1  GI:31876989
KEYWORDS    JP 2002355074-A/25.
SOURCE      Escherichia coli O157:H7
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 14181)
  AUTHORS   Hayashi,H., Shinagawa,H., Makino,K., Hayashi,T., Onishi,S.,
            Hattori,M. and Kurokawa,K.
  TITLE     Nucleic acid molecule and polypeptide specific to intestinal
            hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
  JOURNAL   Patent: JP 2002355074-A 25 10-DEC-2002;
            PRESIDENT OF UNIVERSITY OF TSUKUBA
COMMENT     OS   Escherichia coli O157:H7
            PN   JP 2002355074-A/25
            PD   10-DEC-2002
            PF   24-JAN-2002 JP 2002015959
            PI   HIDEO HAYASHI,HIDEO SHINAGAWA,KOZO MAKINO,TETSUYA HAYASHI,SHIN

            PI   ONISHI,
            PI   MASAHIRA HATTORI,KEN KUROKAWA
            PC   C12N15/09,C12N15/09,A61K31/7088,A61K39/00,A61K48/00,A61P31/04,
            PC   C07K14/245,
            PC   C07K16/12,C12M1/00,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/
            PC   02,C12Q1/68,
            PC   G01N33/15,G01N33/50,G01N33/53,G01N33/53,G01N33/566,G01N37/00,
            PC   C12N15/00,
            PC   C12N15/00,C12N5/00
            CC   Nucleic acid molecule and polypeptide specific to intestinal
            CC           hemorrhagic
            CC   pathogenic Escherichia coli O157:H7, and method of use thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14181
            FT                   /organism='Escherichia coli O157:H7'.
FEATURES             Location/Qualifiers
     source          1. .14181
                     /organism="Escherichia coli O157:H7"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:83334"
ORIGIN

  Query Match             100.0%;  Score 666;  DB 6;  Length 14181;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-120;
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 781

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        780 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 721

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 661

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        660 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 601

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 541

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 481

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 421

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 361

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 301

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 241

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 181

Qy        661 ATTTAA 666
              ||||||
Db        180 ATTTAA 175


RESULT 5
AR204225/c
LOCUS       AR204225               14187 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 121 from patent US 6365723.
ACCESSION   AR204225
VERSION     AR204225.1  GI:21500820
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14187)
  AUTHORS   Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
  TITLE     Sequences of E. coli O157
  JOURNAL   Patent: US 6365723-A 121 02-APR-2002;
FEATURES             Location/Qualifiers
     source          1. .14187
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           100.0%;  Score 666;  DB 6;  Length 14187;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-120;
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 786

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        785 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 726

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        725 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 666

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        665 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 606

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 546

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 486

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 426

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 366

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 306

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 246

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 186

Qy        661 ATTTAA 666
              ||||||
Db        185 ATTTAA 180


RESULT 6
AR637574/c
LOCUS       AR637574               14187 bp    DNA     linear   PAT 20-APR-2005
DEFINITION  Sequence 121 from patent US 6855814.
ACCESSION   AR637574
VERSION     AR637574.1  GI:62771316
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14187)
  AUTHORS   Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
  TITLE     Sequences of E. coli O157
  JOURNAL   Patent: US 6855814-A 121 15-FEB-2005;
            Wisconsin Alumni Research Foundation; Madison, WI
FEATURES             Location/Qualifiers
     source          1. .14187
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match           100.0%;  Score 666;  DB 6;  Length 14187;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-120;
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 786

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        785 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 726

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        725 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 666

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        665 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 606

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 546

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 486

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 426

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 366

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 306

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 246

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 186

Qy        661 ATTTAA 666
                ||||||
Db        185 ATTTAA 180


RESULT 7
AB008676
LOCUS       AB008676               31482 bp    DNA     linear   BCT 19-MAY-1999
DEFINITION  Escherichia coli 0157 DNA, map position at 46 min., complete cds.
ACCESSION   AB008676
VERSION     AB008676.1  GI:4867913
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Shimizu,T., Yamasaki,S., Tsukamoto,T. and Takeda,Y.
  TITLE     Analysis of the genes responsible for the O-antigen synthesis in
            enterohaemorrhagic Escherichia coli O157
  JOURNAL   Microb. Pathog. 26 (5), 235-247 (1999)
   PUBMED   10222209
REFERENCE   2  (bases 1 to 31482)
  AUTHORS   Shimizu,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1997) Takeshi Shimizu, International Medical
            Center of Japan, Department of Appropriate Technology Development,
            Research Institute; 1-21-1 Toyama, Shinjuku-ku, Tokyo 162, Japan
            (E-mail:kshimizu@imcj.go.jp, Tel:+81-3-3202-7181,
            Fax:+81-3-3202-7364)
FEATURES             Location/Qualifiers
     source          1. .31482
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="184"
                     /serotype="O157"
                     /db_xref="taxon:562"
                     /map="46 min."
                     /clone="1-4-1"
     gene            505. .1398
                     /gene="galF"
     CDS             505. .1398
                     /gene="galF"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="UTP-glucose-1-phosphate uridylyltransferase"
                     /protein_id="BAA77723.1"
                     /db_xref="GI:4867914"
                     /translation="MTNLKAVIPVAGLGMHMLPATKAIPKEMLPIVDKPMIQYIVDEI
                     VAAGIKEILLVTHASKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNV
                     RQGEPLGLGHSILCARPAIGDNPFVVVLPDVVIDDASADPLRYNLAAMIARFNETGRS
                     QVLAKRMPGDLSEYSVIQTKEPLDREGKVSRIVEFIEKPDQPQTLDSDIMAVGRYVLS
                     ADIWPELERTQPGAWGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKY
                     GLRNLKEGAKFRKGIEKLLSE"
     gene            1717. .2499
                     /gene="wbhA"
     CDS             1717. .2499
                     /gene="wbhA"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyl transferase"
                     /protein_id="BAA77724.1"
                     /db_xref="GI:4867915"
                     /translation="MNKETVSIIMPVYNGAKTIISSVESIIHQSYQDFVLYIIDDCST
                     DDTFSLINSRYKNNQKIRILRNKTNLGVAESRNYGIEMATGKYISFCDADDLWHEKKL
                     ERQIEVLNNECVDVVCSNYYVIDNKRNIVGEVNAPHVINYRKMLMKNYIGNLTGIYNA
                     NKLGKFYQKKIGHEDYLMWLEIINKTNGAICIQDNLAYYMRSNNSLSGNKIKAAKWTW
                     SIYREHLHLSFPKTLYYFLLYASNGVMKKITHSLLRRKETKK"
     gene            3153. .3680
                     /gene="wbhB"
     CDS             3153. .3680
                     /gene="wbhB"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase"
                     /protein_id="BAA77725.1"
                     /db_xref="GI:4867916"
                     /translation="MLRIINRLTIEQVESVPYVVSESIKNDFFPTPELEKELKAIINR
                     IQGIKHQDLFYGERLHKQVFGDMGANFLSVTTYGAELLVFFGFLCVFIIPLGIYIPFY
                     LLKRMKKTHSSINCAFYSYIIMILLQYLVAGNASAFFFGPFLSVLIMCTPLILLHDTL
                     KRLSRNENISYNCDL"
     gene            3649. .4395
                     /gene="wbhC"
     CDS             3649. .4395
                     /gene="wbhC"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyl transferase"
                     /protein_id="BAA77726.1"
                     /db_xref="GI:4867917"
                     /translation="MKISVITVTYNNAEGLEKTLSSLSILKIKPFEIIIVDGGSTDGT
                     NRVISRFTSMNITHVYEKDEGIYDAMNKGRMLAKGDLIHYLNAGDSVIGDIYKNIKEP
                     CLIKVGLFENDKLLGFSSITHSNTGYCHQGVIFPKNHSEYDLRYKICADYKLIQEVFP
                     EGLRSLSLITSGYVKYDMGGVSSKKRILRDKELAKIMFEKNKKNLIKFIPISIIKILF
                     PERLRRVLRKMQYICLTLFFMKNSSPYDNE"
     gene            4388. .5773
                     /gene="wzx"
     CDS             4388. .5773
                     /gene="wzx"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="flippase"
```

/protein_id="BAA77727.1"
/db_xref="GI:4867918"
/translation="MNKIKKILKFCTLKKYDTSSALGREQERYRIISLSVISSLISKI
LSLLSLILTVSLTLPYLGQERFGVWMTITSLGAALTFLDLGIGNALTNRIAHSFACGK
NLKMSRQISGGLTLLAGLSFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKVFV
IIFGLGIYSNGVQKVYMGIQKAYISNIVNAIFILLSIITLVISSKLHAGLPVLIVSTL
GIQYISGIYLTINLIIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSGDN
FIISITLGVTYVAVFSITQRLFQISTVPLTIYNIPLWAAYADAHARNDTQFIKKTLRT
SLKIVGISSFLLAFILVVFGSEVVNIWTEGKIQVPRTFIIAYALWSVIDAFSNTFASF
LNGLNIVKQQMLAVVTLILIAIPAKYIIVSHFGLTVMLYCFIFIYIVNYFIWYKCSFK
KHIDRQLNIRG"
gene 5776. .6870
/gene="wbhD"
CDS 5776. .6870
/gene="wbhD"
/note="putative"
/codon_start=1
/transl_table=11
/product="perosamine synthetase"
/protein_id="BAA77728.1"
/db_xref="GI:4867919"
/translation="MKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKFAE
QNHVQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFVDS
DNETWQMSVSDIEQKITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEAFG
SKYKGKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHRQY
WHDVIGYNYRMTNICAAIGLAQLEQADDFISRKREIADIYKKNINSLVQVHKESKDVF
HTYWMVSILTRTAEEREELRNHLADKLIETRPVFYPVHTMPMYSEKYQKHPIAEDLGW
RGINLPSFPSLSNEQVIYICESINEFYSDK"
gene 6895. .8109
/gene="wbhE"
CDS 6895. .8109
/gene="wbhE"
/note="putative"
/codon_start=1
/transl_table=11
/product="glycosyl transferase"
/protein_id="BAA77729.1"
/db_xref="GI:4867920"
/translation="MKIALNSDGFYEWGGGIDFIKYILSILETKPEICIDILLPRNDI
HSLIREKAFPFKSILKAILKRERPRWISLNRFNEQYYRDAFTQNNIETNLTFIKSKSS
AFYSYFDSSDCDVILPCMRVPSGNLNKKAWIGYIYDFQHCYYPSFFSKREIDQRNVFF
KLMLNCANNIIVNAHSVITDANKYVGNYSAKLHSLPFSPCPQLKWFADYSGNIAKYNI
DKDYFIICNQFWKHKDHATAFRAFKIYTEYNPDVYLVCTGATQDYRFPGYFNELMVLA
KKLGIESKIKILGHIPKLEQIELIKNCIAVIQPTLFEGGPGGGVTFDAIALGKKVILS
DIDVNKEVNCGDVYFFQAKNHYSLNDAMVKADESKIFYEPTTLIELGLKRRNACADFL
LDVVKQEIESRS"
gene 8129. .9247
/gene="gmd"
CDS 8129. .9247
/gene="gmd"
/note="putative"
/codon_start=1
/transl_table=11
/product="GDP-mannose 4,6-dehydratase"
/protein_id="BAA77730.1"
/db_xref="GI:4867921"
/translation="MTKVALITGVTGQDGSYLAEFLLDKGYEVHGIKRRASSFNTERI
DHIYQDPHGSNPNFHLHYGDLTDSSNLTRILKEVQPDEVYNLAAMSHVAVSFESPEYT
ADVDAIGTLRLLEAIRFLGLENKTRFYQASTSELYGLVQEIPQKESTPFYPRSPYAVA
KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRGLANIAQGLESCLYL
GNMDSLRDWGHAKDYVRMQWLMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLGIKMSF
VGKGIEEKGIVDSVEGQDAPGVKPGDVIVAVDPRYFRPAEVDTLLGDPSKANLKLGWR
PEITLAEMISEMVAKDLEAAKKHSLLKSHGFSVSLALE"
gene 9250. .10215
/gene="wbhF"
CDS 9250. .10215
/gene="wbhF"
/note="putative"
/codon_start=1
/transl_table=11
/product="GDP-L-fucose pathway enzyme"
/protein_id="BAA77731.1"
/db_xref="GI:4867922"
/translation="MNKQRIFIAGHQGMVGSAITRRLKQRDDVELVLRTRDELNLLDS
SAVLDFFSSQKIDQVYLAAAKVGGILANSSYPADFIYENIMIEANVIHAAHKNNVNKL
LFLGSSCIYPKLAHQPIMEDELLQGKLEPTNEPYAIAKIAGIKLCESYNRQFGRDYRS
VMPTNLYGPNDNFHPSNSHVIPALLRRFHDAVENNSPNVVVWGSGTPKREFLHVDDMA
SASIYVMEMPYDIWQKNTKVMLSHINIGTGIDCTICELAETIAKVVGYKGHITFDTTK
PDGAPRKLLDVTLLHQLGWNHKITLHKGLENTYNWFLENQLQYRG"
gene 10218. .10727
/gene="wbhG"
CDS 10218. .10727
/gene="wbhG"
/note="putative"
/codon_start=1
/transl_table=11
/product="GDP-L-fucose pathway enzyme"
/protein_id="BAA77732.1"
/db_xref="GI:4867923"
/translation="MFLHSQDFATIVRSTPLISIDLIVENEFGEILLGKRINRPAQGY
WFVPGGRVLKDEKLQTAFERLTEIELGIRLPLSVGKFYGIWQHFYEDNSMGGDFSTHY
IVIAFLLKLQPNILKLPKSQHNAYCWLSRAKLINDDDVHYNCRAYFNNKTNDAIGLDN
KDIICLMRQ"
gene 10739. .12157
/gene="manC"
CDS 10739. .12157
/gene="manC"
/note="putative"
/codon_start=1
/transl_table=11

Query Match 100.0%; Score 666; DB 1; Length 31482;
Best Local Similarity 100.0%; Pred. No. 4.2e-120;
Matches 666; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14794 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 14853

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14854 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 14913

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14914 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 14973

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14974 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 15033

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15034 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 15093

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15094 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 15153

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15154 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 15213

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15214 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 15273

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15274 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 15333

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15334 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 15393

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15394 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 15453

Qy        661 ATTTAA 666
              ||||||
Db      15454 ATTTAA 15459


RESULT 8
AE005174_28/c
WPCOMMENT
  Sequence split into 56 fragments  LOCUS AE005174 Accession AE005174
     Fragment Name      Begin     End
     AE005174_00            1    110000
     AE005174_01       100001    210000
     AE005174_02       200001    310000
     AE005174_03       300001    410000
     AE005174_04       400001    510000
     AE005174_05       500001    610000
     AE005174_06       600001    710000
     AE005174_07       700001    810000
     AE005174_08       800001    910000
     AE005174_09       900001   1010000
     AE005174_10      1000001   1110000
     AE005174_11      1100001   1210000
     AE005174_12      1200001   1310000
     AE005174_13      1300001   1410000
     AE005174_14      1400001   1510000
     AE005174_15      1500001   1610000
     AE005174_16      1600001   1710000
     AE005174_17      1700001   1810000
     AE005174_18      1800001   1910000
     AE005174_19      1900001   2010000
     AE005174_20      2000001   2110000
     AE005174_21      2100001   2210000
     AE005174_22      2200001   2310000
     AE005174_23      2300001   2410000
     AE005174_24      2400001   2510000
     AE005174_25      2500001   2610000
     AE005174_26      2600001   2710000
     AE005174_27      2700001   2810000
     AE005174_28      2800001   2910000
     AE005174_29      2900001   3010000
     AE005174_30      3000001   3110000
     AE005174_31      3100001   3210000
     AE005174_32      3200001   3310000
     AE005174_33      3300001   3410000
     AE005174_34      3400001   3510000
     AE005174_35      3500001   3610000
     AE005174_36      3600001   3710000
     AE005174_37      3700001   3810000
     AE005174_38      3800001   3910000
     AE005174_39      3900001   4010000
     AE005174_40      4000001   4110000
     AE005174_41      4100001   4210000
     AE005174_42      4200001   4310000
     AE005174_43      4300001   4410000
     AE005174_44      4400001   4510000
     AE005174_45      4500001   4610000
     AE005174_46      4600001   4710000
     AE005174_47      4700001   4810000
     AE005174_48      4800001   4910000
     AE005174_49      4900001   5010000
     AE005174_50      5000001   5110000
     AE005174_51      5100001   5210000
     AE005174_52      5200001   5310000
     AE005174_53      5300001   5410000
     AE005174_54      5400001   5510000
     AE005174_55      5500001   5528445
Continuation (29 of 56) of AE005174 from base 2800001 (AE005174 Escherichia coli O157:H7

  Query Match             100.0%;  Score 666;  DB 1;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-120;
```

```
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44410 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 44351

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44350 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 44291

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44290 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 44231

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44230 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 44171

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44170 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 44111

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44110 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 44051

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44050 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 43991

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43990 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 43931

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43930 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 43871

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43870 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 43811

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43810 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 43751

Qy        661 ATTTAA 666
              ||||||
Db      43750 ATTTAA 43745


RESULT 9
BA000007_27/c
WPCOMMENT
  Sequence split into 55 fragments  LOCUS BA000007 Accession BA000007
     Fragment Name      Begin     End
     BA000007_00            1    110000
     BA000007_01       100001    210000
     BA000007_02       200001    310000
     BA000007_03       300001    410000
     BA000007_04       400001    510000
     BA000007_05       500001    610000
     BA000007_06       600001    710000
     BA000007_07       700001    810000
     BA000007_08       800001    910000
     BA000007_09       900001   1010000
     BA000007_10      1000001   1110000
     BA000007_11      1100001   1210000
     BA000007_12      1200001   1310000
     BA000007_13      1300001   1410000
     BA000007_14      1400001   1510000
     BA000007_15      1500001   1610000
     BA000007_16      1600001   1710000
```

```
     BA000007_17       1700001   1810000
     BA000007_18       1800001   1910000
     BA000007_19       1900001   2010000
     BA000007_20       2000001   2110000
     BA000007_21       2100001   2210000
     BA000007_22       2200001   2310000
     BA000007_23       2300001   2410000
     BA000007_24       2400001   2510000
     BA000007_25       2500001   2610000
     BA000007_26       2600001   2710000
     BA000007_27       2700001   2810000
     BA000007_28       2800001   2910000
     BA000007_29       2900001   3010000
     BA000007_30       3000001   3110000
     BA000007_31       3100001   3210000
     BA000007_32       3200001   3310000
     BA000007_33       3300001   3410000
     BA000007_34       3400001   3510000
     BA000007_35       3500001   3610000
     BA000007_36       3600001   3710000
     BA000007_37       3700001   3810000
     BA000007_38       3800001   3910000
     BA000007_39       3900001   4010000
     BA000007_40       4000001   4110000
     BA000007_41       4100001   4210000
     BA000007_42       4200001   4310000
     BA000007_43       4300001   4410000
     BA000007_44       4400001   4510000
     BA000007_45       4500001   4610000
     BA000007_46       4600001   4710000
     BA000007_47       4700001   4810000
     BA000007_48       4800001   4910000
     BA000007_49       4900001   5010000
     BA000007_50       5000001   5110000
     BA000007_51       5100001   5210000
     BA000007_52       5200001   5310000
     BA000007_53       5300001   5410000
     BA000007_54       5400001   5498450
Continuation (28 of 55) of BA000007 from base 2700001 (BA000007 Escherichia coli O157:H7

  Query Match             100.0%;  Score 666;  DB 1;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-120;
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74197 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 74138

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74137 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 74078

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74077 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 74018

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74017 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 73958

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73957 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 73898

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73897 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 73838

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73837 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 73778

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73777 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 73718

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73717 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 73658

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73657 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 73598

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73597 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 73538

Qy        661 ATTTAA 666
              ||||||
Db      73537 ATTTAA 73532


RESULT 10
AR318202
LOCUS       AR318202                687 bp    DNA     linear   PAT 17-AUG-2003
DEFINITION  Sequence 752 from patent US 6562958.
ACCESSION   AR318202
VERSION     AR318202.1  GI:33699305
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   Breton,G. and Bush,D.
  TITLE     Nucleic acid and amino acid sequences relating to Acinetobacter
            baumannii for diagnostics and therapeutics
  JOURNAL   Patent: US 6562958-A 752 13-MAY-2003;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             32.8%;  Score 218.2;  DB 6;  Length 687;
  Best Local Similarity   60.1%;  Pred. No. 1.8e-32;
  Matches  383;  Conservative    0;  Mismatches  248;  Indels    6;  Gaps    1;

Qy         12 TGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAAT 71
              |||| ||| ||||||  |  ||| |||||  |||   || ||  |  |  ||| ||  |
Db         51 TGGTGTTTATGGTGCCAGTGGTTTTGGTAAGGAAGTTATGCCTTTAGTGCGTCAGCAGTT 110

Qy         72 AAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAA 131
                 |  ||        | | |||       |||| | |||||||   |  | ||||
Db        111 TCCAACATTAAGTAAAGAACAATTTGC------TTTTATTGATGATGGACTATCTGGAAC 164

Qy        132 GAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAA 191
              ||   | ||||| |    ||| || ||       |  |||| | |    |||  |    |
Db        165 GACTTTAAATGGATACCCAGTACTATCCTATCTCGACTTTATCAGTAAACCTGCTGATCA 224

Qy        192 AAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAAT 251
              |||  | | |   | ||||| || |||  |  | | ||  |||   |   |       |
Db        225 TAAGGCTGTGACAATCGCTATAGCAAATAGTGTGGTTCGTGAGAAGTTAGTTAGCCTTTT 284

Qy        252 ATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCA 311
              |  |  | | || ||  |||   || || |  ||  | ||||||  ||  ||  |||| |
Db        285 AGAAAAAGATGGTGTACAACACCTAGCTGTTCAATCTACAAATACAGTAATTCTTGATGA 344

Qy        312 TACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCA 371
                |  |||||||   ||      |  ||| |||||||  |    || ||  ||||||   |
Db        345 GGTTGAGATAGGTGAAGGAAGCTTACTTTGTCCCTTTACTTGCTTAACTTCTAATATAAA 404

Qy        372 TATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGG 431
              || || | |||||| || || || || ||    || || || |||||||||  ||||||
Db        405 AATTGGAAAGTTTTTCCACGCTAATATTTATAGTTATGTGGCTCATGATTGTGTAATAGG 464
```

```
Qy        432 AGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAA 491
              ||| |||||||||||| || ||||| ||||| ||||| ||  || ||  |||||||||  |
Db        465 AGATTATGTTACATTCGCGCCTGGCGCTAAGTGTAACGGTAATATTCATATTGAAGATCA 524

Qy        492 TGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTAT 551
              ||| || || ||| | || ||||| ||||| |||||     ||| ||    ||  | || ||
Db        525 TGCCTACATTGGCACTGGAGCAGTTATTAAACAGGGCACACCTGATAAGCCTTTAATAAT 584

Qy        552 TGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTTCCTGCCGGTAT 611
              |||     || || ||| | || ||||| ||||| || || |||||||| || || ||  |
Db        585 TGGTAAAGGTGCTATTGTTGGAATGGGCGCTGTAGTAACGAAAAGTGTGCCAGCTGGCGT 644

Qy        612 AACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGA 648
              ||||||    || ||||| ||||||    |  |||||
Db        645 AACTGTTGTAGGTAATCCTGCAAGAATTTTAGAAAGA 681
```

RESULT 11
ALW243431
LOCUS ALW243431 26953 bp DNA linear BCT 15-APR-2005
DEFINITION Acinetobacter lwoffii wzc, wzb, wza, weeA, weeB, wceC, wzx, wzy, weeD, weeE, weeF, weeG, weeH, weeI, weeJ, weeK, galU, ugd, pgi, galE, pgm (partial) and mip (partial) genes (emulsan biosynthetic gene cluster), strain RAG-1.
ACCESSION AJ243431
VERSION AJ243431.1 GI:6009984
KEYWORDS acetyltransferase; amino-transferase; dehydrogenase; dTDP-glucose-4,6-dehydratase; emulsan repeating unit flippase; emulsan repeating unit polymerase; galactoside acetyltransferase; galE gene; galU gene; glycosyl transferase; macrophage infectivity potentiator; MIP gene; NDP-N-Acetyl-D-Galactosaminuronic acciddehydrogenase; outer membrane protein; pgi gene; pgm gene; phosphoglucomutase; phosphoglucose isomerase; protein tyrosine kinase; protein tyrosine phosphatase; UDP-galactose phosphate transferase; UDP-glucose 4-epimerase; UDP-glucose dehydrogenase; UDP-N-acetylglucosamine 2-epimerase; ugd gene; UTP-glucose-1-phosphate uridylyltransferase; weeA gene; weeB gene; weeC gene; weeD gene; weeE gene; weeF gene; weeG gene; weeH gene; weeI gene; weeJ gene; weeK gene; wza gene; wzb gene; wzc gene; wzx gene; wzy gene.
SOURCE Acinetobacter lwoffii
ORGANISM Acinetobacter lwoffii
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter.
REFERENCE 1
AUTHORS Nakar,D. and Gutnick,D.L.
TITLE A protein tyrosine kinase of Acinetobacter lwoffii RAG-1 is involved in emulsan biosynthesis
JOURNAL Unpublished
REFERENCE 2
AUTHORS Nakar,D. and Gutnick,D.L.
TITLE Genomic organization of the wce region of Acinetobacter lwoffii RAG-1 required for emulsan biosynthesis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 26953)
AUTHORS Gutnick,D.L.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-1999) Gutnick D.L., Molecular Microbiology and Biotechnology, The George S. Wise Faculty of Life Sciences, Tel-Aviv University, Ramat Aviv, 69978, ISRAEL
REMARK Revised by author 19-JUL-1999
FEATURES Location/Qualifiers

```
     source          1. .26953
                     /organism="Acinetobacter lwoffii"
                     /mol_type="genomic DNA"
                     /strain="RAG-1"
                     /db_xref="taxon:28090"
                     /clone_lib="EcoR1 mini-library"
                     /country="Israel:Tel-Aviv"
     gene            complement(<1. .534)
                     /gene="mip"
     CDS             complement(<1. .534)
                     /gene="mip"
                     /codon_start=1
                     /transl_table=11
                     /product="putative macrophage infectivity potentiator"
                     /protein_id="CAB57192.1"
                     /db_xref="GI:6009985"
                     /db_xref="GOA:Q9RME0"
                     /db_xref="InterPro:IPR000774"
                     /db_xref="InterPro:IPR001179"
                     /db_xref="UniProt/TrEMBL:Q9RME0"
                     /translation="MSKALPFVVAAVLGGAALVPVYYATQHPTTELGAKADKNAAPVA
                     KISYALGYEVAKQTPPELDTNSFITGFRDAHAQKPAAYTEEELQAAYVDFQKEIQQKQ
                     LNETKKAEATSNSFLAENAKKEGVKTTASGLQYKIIKEGTGKQPSATSVVKVHYKGQL
                     TDGKVFDSSYDRGEPIEF"
     gene            complement(711. .2891)
                     /gene="wzc"
     CDS             complement(711. .2891)
                     /gene="wzc"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="protein tyrosine kinase"
                     /protein_id="CAB57193.1"
                     /db_xref="GI:6009986"
                     /db_xref="GOA:Q9RMD9"
                     /db_xref="InterPro:IPR003856"
                     /db_xref="UniProt/TrEMBL:Q9RMD9"
                     /translation="MSQNTNTEDTIDLKELFFALIAQWKLIALCIILSLVFALLYLRA
                     TPDTYSVDALVQVEDSKGASAALLGDLSQMIEQKSPAQSEIEILKSRLVLGSVIQALH
                     LDLEIASENNTFSHRLLNKTDYETEYSPKTVTFNDGEKFFNINQFDVPNSYLDKKLSL
                     SFKNNNFRLVDPKTDEIVLTAPLNKLNQLKNQFGTWKVAIFTNDTFDANYFVQKNSLP
                     AAVQAILSNYSVAERGKLTGILGLNYQGQDKEHITKVLNAILAAYSAQNVERRSAETA
                     QTLKFLDEQLPDLKKQLDDAERVFNKFRQQYNTVDVTKESELYLTQSITLETKKIELQ
                     QQQADLAAKYTAEHPAIREINAQITALDKQITQLNSTLKQLPDIQRQYLQLFREVEVK
                     TQLYTALLNSYQQLRIAKAGEIGNVRIVDTAVEPVEPIKPKKLQILILSIFLGGFLGT
                     LLALLRNMLRTGVKDSTQIEREFDLPVYATVPRSPVQETRMSILKKKKSIPILAVKHS
                     DDIAIESLRSIRTAIHFSLASAKNNIIMIAGPAPEVGKSFISTNLSTIFAQSDKKVLL
                     IDADMRRGYMHKYFDVDIKPGLSEFLSNQAELSQVIHQTEVNGLNVITRGKSPTNPSE
                     MLGSQKFKALLDELSQQYDHIIIDTPPILAVTDGIIISQYAGVNLLIARYSKSQMKEL
                     ELSLNRFEQAGVKVNGFILNDIQRTAGAGYSYNYAYAYKANKED"
     gene            complement(2911. .3339)
                     /gene="wzb"
     CDS             complement(2911. .3339)
                     /gene="wzb"
                     /codon_start=1
                     /transl_table=11
                     /product="putative protein tyrosine phosphatase"
                     /protein_id="CAB57194.1"
                     /db_xref="GI:6009987"
                     /db_xref="GOA:Q9RMD8"
                     /db_xref="InterPro:IPR000106"
                     /db_xref="UniProt/TrEMBL:Q9RMD8"
                     /translation="MNIKNVLVVCVGNICRSPMAEYLLKQQHPNLRIESAGISGMIGH
                     AADDKAQLCMRRLDIDMSTHIAKKLNSEHLKQADLILVMSNNQQKHIEQTWPFAKGKV
                     FRLGHWQNKNVPDPYQHDQAFFDQTCLLIQQCVTDWKNYI"
     gene            complement(3345. .4445)
                     /gene="wza"
     CDS             complement(3345. .4445)
                     /gene="wza"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="CAB57195.1"
                     /db_xref="GI:6009988"
                     /db_xref="GOA:Q9RMD7"
                     /db_xref="InterPro:IPR003715"
                     /db_xref="UniProt/TrEMBL:Q9RMD7"
                     /translation="MKHYQLFSILAISISATGCAVTSGLQTYDIPSEGIYTTDQGTDV
                     NVIKITQDNLPAIQPAQIDYQSNYSTLFKQQQQAYRLSSGDVLSIQLWAYPEITPPVN
                     NLTNDQSIQANGYPIDYNGFVQIPLVGRYKASGKTLTQLNTDLRNQFARYLKNPDVIV
                     RVLSYQGQRYSVQGSVLKGGQFFLNDQPISIYSALGMAGGVTEKGDNTFIQLIRNGIT
                     YDLNTIALEKAGYSLNKLLVQPNDTIYVSTKENQKIYVMGESGKNQALPMRDQGMTLS
                     DALGESLGLNPLSASASRIYVLRTNPNDRHTELYHLNLMSLGDFGLANQFRLRSNDIV
```

```
                YVDATGLTRWQRVINQIIPFSSALYNIDRLGQ"
gene            5062. .6192
                /gene="weeA"
CDS             5062. .6192
                /gene="weeA"
                /codon_start=1
                /transl_table=11
                /product="putative UDP-N-acetylglucosamine 2-epimerase"
                /protein_id="CAB57196.1"
                /db_xref="GI:6009989"
                /db_xref="GOA:Q9RMD6"
                /db_xref="InterPro:IPR003331"
                /db_xref="UniProt/TrEMBL:Q9RMD6"
                /translation="MKKVLLVFGTRPEAIKMAPLALKLQQFNEDFETKVCVTGQHRQM
                LDQVLSLFELTPDFDLNLMKPGQTLSDVTSGVLKGLEQVFAEWMPDLILVHGDTATTF
                AASLAGYYHKIKIGHVEAGLRTGDLYSPWPEEANRQLTGVLANFHFAPTRSSYNNLIK
                ENVNPETVVITGNTVIDALLQVKTKVEQEQQLIQKFEQEFSFIDKSKKLILVTGHRRE
                NFGQGFLNICTALANIAKKYPDIQIVYPVHLNPNVQQPVNELLANIDNVFLIAPQDYL
                PFVYLMNRSYLILTDSGGIQEEAPSLGKPVLVMRDTTERPEAVDAGTVRLVGTDVNAI
                QSAVSELLENTTIYASMAEAHNPYGDGTACQRIIDFLKVNFI"
gene            6225. .7478
                /gene="weeB"
CDS             6225. .7478
                /gene="weeB"
                /codon_start=1
                /transl_table=11
                /product="putative NDP-N-acetyl-D-galactosaminuronic acid
                dehydrogenase"
                /protein_id="CAB57197.1"
                /db_xref="GI:6009990"
                /db_xref="GOA:Q9RMD5"
                /db_xref="InterPro:IPR000169"
                /db_xref="InterPro:IPR000205"
                /db_xref="InterPro:IPR001732"
                /db_xref="UniProt/TrEMBL:Q9RMD5"
                /translation="MSSSFKHICVIGLGYIGLPTAATFAAHGVKVTGVDVNQHAVDMI
                NQGKVHIVEPDLDALVRDVVAQGKLSAQLAPTEADAYIVAVPTPFKDDYQPDLKYIEA
                ASKALAPFLKSGNLVILESTSPVEATEQMSAWLSEARPDLSFPQQKGEEADILVAHCP
                ERVLPGKVLQELISNDRIIGGMTPRCSKAACDLYKIFVKGACIETNARTAEMCKLTEN
                SFRDVNIAFANELSVICDKLDINVWELISLANRHPRVNILQPGPGVGGHCIAVDPWFI
                VAKTPEQARLIRTAREVNDAKPEWVIDQVKIKIADFLQANPTKTIQDVTIACYGLAFK
                PDIDDLRESPALEITKKLAEQGFNILAIEPNIHELPAKLPSNIKLIALKEHQQADIHL
                ILVDHRQFKNMNFSFSYVVDTKGIL"
gene            7479. .8033
                /gene="weeC"
CDS             7479. .8033
                /gene="weeC"
                /codon_start=1
                /transl_table=11
                /product="putative galactoside acetyltransferase"
                /protein_id="CAB57198.1"
                /db_xref="GI:6009991"
                /db_xref="GOA:Q9RMD4"
                /db_xref="UniProt/TrEMBL:Q9RMD4"
                /translation="MYSLFKFFFRLGDYYRGLKLFLLINLCGGICKGIPKVGKNVIFK
                YPPHKGIKFGKKCDIGAFSHFDIPPFGQLIVGDCVKMTQGVVVSVINKVEIQSNTLIA
                EWVSIRDAQHLFDKNEPINKQGMKKGEILIEEDVWIGRGSSIFLNTTIRKGSIIGAES
                IVKSIEISSMEIHAGNPLKFIKNR"
gene            8039. .9244
                /gene="wzx"
CDS             8039. .9244
                /gene="wzx"
                /codon_start=1
                /transl_table=11
                /product="putative emulsan repeating unit flippase"
                /protein_id="CAB57199.1"
                /db_xref="GI:6009992"
                /db_xref="UniProt/TrEMBL:Q9RMD3"
                /translation="MKILSKLIAGGVIVQILVAVSVMISSRLYKPEVFGELGWYLSFA
                SVFAIVAAMRFDYIVLSDRVCAEEKNEFFSNAIVVSIITFFLLCIGVVIANNVFHFKY
                NIIYLLLFCLSLSFFNLLSQYLILIKDYNKFIFYKTNQLLLQIVMIVGFYYVFHENGL
                LLANILPQLIISILFFLFIRKSLLFKLNNFIPFFYKNYREGFKNSFLTFIQYSTPVIP
                VLFGGYVFDKSQIGAYFLFSQTMAVPFSLIRRNVLILFNGEYSDAKKIKELFESILTT
```

```
                RNFIIVLMSFACFLVFLVFCSSYIIMIFFGREWLEFSWLLSLLFIYYFFDALLQPLTT

  Query Match           27.6%;  Score 183.6;  DB 1;  Length 26953;
  Best Local Similarity  58.4%;  Pred. No. 3.3e-26;
  Matches  320;  Conservative    0;  Mismatches  228;  Indels    0;  Gaps    0;

Qy        101 TGGTTTTTGTGGATGATGTTTTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAA 160
              | |||||| | ||||||| |||||  | | | ||||||| |  || | |
Db      17375 TTGTTTTTATTGATGATGGCCAAGCAGGTGGTAGACTGAATGGTTATTCGGTTTTAACCT 17434

Qy        161 CCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATG 220
              ||    ||| | ||  | |  |  | | ||| | | |  |||| || || ||
Db      17435 ACACAGATTTTATTAATAACACAGCTACACATAAGGCTGTAACGATTGCGATCGCGAACA 17494

Qy        221 ATAAGATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTA 280
              ||  | ||  | | | |  | |  ||       | || ||| |||  || | |
Db      17495 GTATTGTTCGTGAAAAACTAGTCGTGCGTCTAAATGAGGATGGAGTTCAACATTTAGCGA 17554

Qy        281 TAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTT 340
              || ||     ||||  | ||||  ||| |     ||| |||  ||  | |  |||
Db      17555 TACAAGCAGACAATACAATGGTTTTGGATGAGGTAGAGATCGGTGAAGGGAGTCTGCTTT 17614

Qy        341 CTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATAT 400
              || ||   |   |  |  | |||| | | || | | |||||| || || || || |
Db      17615 GCCCATTCACTTGTTTGACCTCAAATATTAAAATTGGCAAGTTTTTCCACGCTAATATCT 17674

Qy        401 ACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTA 460
              |   ||||| || |||||||||   || || || || ||||| ||||| || || |  |
Db      17675 ATAGTTACGTCGCTCATGATTGTGTGATTGGTGATTACGTTACCTTTGCACCAGGAGTAA 17734

Qy        461 AATGTAATGGATATGTTGTTATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTA 520
              |||||||||| || ||  |||||| ||| |||| ||||| ||  | |||||||| || |
Db      17735 AATGTAATGGAAATATTCATATTGAGGACCATGCCTATATTGGTACAGGTGCAGTGATCA 17794

Qy        521 AGCAGGGTGTTCCTAATCGCCCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGG 580
              | || ||    ||  |    ||  |  | |||||    || ||  |  | || ||||| |
Db      17795 AACAAGGAACACCAGACAAGCCCTTGGTGATTGGTAAAGGTGCTGTAGTTGGGATGGGTG 17854

Qy        581 CTGTTGTCACTAAAAGTGTTCCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAA 640
              | || || ||||| ||||| || || ||  ||||||||   || || |||||||||
Db      17855 CAGTGGTAACTAAGAGTGTGCCAGCAGGCGTAACTGTGATTGGTAACCCAGCAAGAATTT 17914

Qy        641 TGAAAAGA 648
              |  ||| |
Db      17915 TAGAAAAA 17922
```

```
RESULT 12
CR543861_00
WPCOMMENT
  Sequence split into 36 fragments  LOCUS CR543861 Accession CR543861
     Fragment Name      Begin     End
     CR543861_00            1    110000
     CR543861_01       100001    210000
     CR543861_02       200001    310000
     CR543861_03       300001    410000
     CR543861_04       400001    510000
     CR543861_05       500001    610000
     CR543861_06       600001    710000
     CR543861_07       700001    810000
     CR543861_08       800001    910000
     CR543861_09       900001   1010000
     CR543861_10      1000001   1110000
     CR543861_11      1100001   1210000
     CR543861_12      1200001   1310000
     CR543861_13      1300001   1410000
     CR543861_14      1400001   1510000
     CR543861_15      1500001   1610000
     CR543861_16      1600001   1710000
     CR543861_17      1700001   1810000
     CR543861_18      1800001   1910000
     CR543861_19      1900001   2010000
     CR543861_20      2000001   2110000
```

```
     CR543861_21       2100001   2210000
     CR543861_22       2200001   2310000
     CR543861_23       2300001   2410000
     CR543861_24       2400001   2510000
     CR543861_25       2500001   2610000
     CR543861_26       2600001   2710000
     CR543861_27       2700001   2810000
     CR543861_28       2800001   2910000
     CR543861_29       2900001   3010000
     CR543861_30       3000001   3110000
     CR543861_31       3100001   3210000
     CR543861_32       3200001   3310000
     CR543861_33       3300001   3410000
     CR543861_34       3400001   3510000
     CR543861_35       3500001   3598621
LOCUS       CR543861               3598621 bp    DNA     circular BCT 17-APR-2005
DEFINITION  Acinetobacter sp. ADP1 complete genome.
ACCESSION   CR543861
VERSION     CR543861.1  GI:49529273
KEYWORDS    complete genome.
SOURCE      Acinetobacter sp. ADP1
  ORGANISM  Acinetobacter sp. ADP1
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Moraxellaceae; Acinetobacter.
REFERENCE   1  (bases 1 to 3598621)
  AUTHORS   Barbe,V., Vallenet,D., Fonknechten,N., Kreimeyer,A., Oztas,S.,
            Labarre,L., Cruveiller,S., Robert,C., Duprat,S., Wincker,P.,
            Ornston,L.N., Weissenbach,J., Marliere,P., Cohen,G.N. and
            Medigue,C.
  TITLE     Unique features revealed by the genome sequence of Acinetobacter
            sp. ADP1, a versatile and naturally transformation competent
            bacterium
  JOURNAL   (er) Nucleic Acids Res. 32 (19), 5766-5779 (2004)
   PUBMED   15514110
REFERENCE   2  (bases 1 to 3598621)
  AUTHORS   Barbe,V., Vallenet,D., Fonknechten,N., Kreimeyer,A., Oztas,S.,
            Labarre,L., Cruveiller,S., Robert,C., Duprat,S., Wincker,P.,
            Ornston,L.N., Weissenbach,J., Marliere,P., Cohen,G.N. and
            Medigue,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Annotation data relative to COG assignations, enzymatic function
            prediction (PRIAM software), TMHMM and SignalP predictions, and
            synteny results (Syntonizer software) are available in
            Acinetobacter database
            (http://www.genoscope.cns.fr/agc/mage/acinetopublic). See this web
            site
            (http://www.genoscope.cns.fr/externe/English/Projets/Projet_DY) for
            more information on the overall project.
            Each annotation includes a confidence level as follow: 1  :
            Function experimentally demonstrated in the studied organism 2a :
            Function of homologous gene experimentally demonstrated in an other
                 organism
            2b : Function of strongly homologous gene
            3  : Function proposed based on presence of conserved amino acid
            motif,
                 structural feature or limited homolgy
            4  : Homolgs of previously reported genes of unknown function 5  :
            No homology to any previously reported sequences 6  : Doubtful CDS
            7  : Gene remnant
            Classification of gene products by their type and by their cellular
            role(s) was made with the MultiFun system
            (http://genprotec.mbl.edu/).
FEATURES             Location/Qualifiers
     source          1. .3598621
                     /organism="Acinetobacter sp. ADP1"
                     /mol_type="genomic DNA"
                     /strain="ADP1"
                     /db_xref="taxon:62977"
     gene            201. .1598
                     /gene="dnaA"
                     /locus_tag="ACIAD0001"
     CDS             201. .1598
                     /gene="dnaA"
                     /locus_tag="ACIAD0001"
                     /function="DNA replication (MultiFun:2.1.1)"
                     /function="action unknown (MultiFun:3.1.2.5)"
                     /function="nucleoproteins, basic proteins
                     (MultiFun:2.3.7)"
                     /function="regulon (MultiFun:3.3.2)"
                     /note="Evidence 2a : Function of homologous gene
                     experimentally demonstrated in an other organism; Product
                     type f : factor"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="DNA replication initiator protein,
                     transcriptional regulator of replication and housekeeping
                     genes"
                     /protein_id="CAG66986.1"
                     /db_xref="GI:49529274"
                     /db_xref="GOA:Q6FG21"
                     /db_xref="InterPro:IPR001957"
                     /db_xref="InterPro:IPR003593"
                     /db_xref="UniProt/Swiss-Prot:Q6FG21"
                     /translation="MLWTDCLTRLRQELSDNVFAMWIRPLVAEETTDSLRLYAPNPYW
                     TRYIQEHHLELISILVEQLSEGRIRQVEILVDSRPGAILSPAEQPATTTAALSSTPVV
                     PQRVKKEVVEPAATQSNKILNSKKRLLNPLFTFSLFVEGRSNQMAAETCRKVLTQLGA
                     SQHNPLFLYGPTGLGKTHLMQAVGNALLQAKPNARVMYMTAESFVQDFVSSLQKGKVE
                     EFKKNCRSLDLLLVDDIHLLAGKEASLVEFFYTFNALLDESKQIILTSDRYPKELTEL
                     DPRLVSRFSWGLSVGVEPPDIETRIEILLKKAENSGVDLPRNCALFIAQQVVANVREL
                     EGALNKVVAIARFKGSQIDLDVVRESLKDVLAIRARTISVENIQRVVSEYFRIPLKEL
                     IGPKRTRIYARPRQLAMGLARELTGDSFPEIGMAFGGRDHSTVMHACEKVQSLKQEDP
                     IFNEDYKNLLRLLQS"
     gene            1834. .2982
                     /gene="dnaN"
                     /locus_tag="ACIAD0002"
     CDS             1834. .2982
                     /gene="dnaN"
                     /locus_tag="ACIAD0002"
                     /EC_number="2.7.7.7"
                     /function="DNA replication (MultiFun:2.1.1)"
                     /note="Evidence 2a : Function of homologous gene
                     experimentally demonstrated in an other organism; Product
                     type e : enzyme"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="DNA polymerase III, beta chain"
                     /protein_id="CAG66987.1"
                     /db_xref="GI:49529275"
                     /db_xref="GOA:Q6FG20"
                     /db_xref="InterPro:IPR001001"
                     /db_xref="UniProt/TrEMBL:Q6FG20"
                     /translation="MRLKIAKESLLNVLSHVVGAVERRHTLNILSNVKIQANAQALTI
                     TGSDLEVELVASTTLAEGACIEAGETTVPARKLVDICKSLPSAALIDLQITEDQRCIL
                     KSGNSRFVLGTLPAEDYPLLTTESSQGTQVQVTQRELKRLFEKTSFAMAVQDVRFYLT
                     GTLLEIDQNQLRAVTTDGHRLALCEVQASSTAMQAVQAIVPRKAVGELQRLLSIEDDQ
                     LSLLIGRELLNVTINIANRDKEQHPITVRFTTKLIDGKFPDYRRVIPRGGDKHVQIAH
                     DVFKQSLQRVAILSNEKLRGVFLNFNPDVLQLRANNPEQDEAIEDIAIQYQDASLEMS
                     FNAQYLLDVLSVLDGDDVSMSMTEANQSVLVQDAAHPDQTYVVMPMRV"
     gene            2998. .4074
                     /gene="recF"
                     /locus_tag="ACIAD0003"
     CDS             2998. .4074
                     /gene="recF"
                     /locus_tag="ACIAD0003"
                     /function="DNA recombination (MultiFun:2.1.3)"
                     /function="DNA repair (MultiFun:2.1.4)"
                     /function="DNA replication (MultiFun:2.1.1)"
                     /note="Evidence 2a : Function of homologous gene
                     experimentally demonstrated in an other organism; Product
                     type e : enzyme"
                     /codon_start=1
```

```
                     /evidence=not_experimental
                     /transl_table=11
                     /product="DNA replication, recombinaison and repair
                     protein"
                     /protein_id="CAG66988.1"
                     /db_xref="GI:49529276"
                     /db_xref="GOA:Q6FG19"
                     /db_xref="InterPro:IPR001238"
                     /db_xref="InterPro:IPR003395"
                     /db_xref="InterPro:IPR003439"
                     /db_xref="UniProt/TrEMBL:Q6FG19"
                     /translation="MQITRLNIERVRNLKAVALSGLQPFNIFYGANGSGKTSILEAVH
                     LLATGRSFRTHMPKHYIQQNAQDAIIFAQSLSEKIGMQKLLSGEQLIKVNGDTVATQG
                     QLAKLLPLQHLDPQSTDIIDHGAKPRRQLLDWLMFHVEPEFYFAWQYYSRALKQRNML
                     LKTKRQLSLAELEPWNKMLSEYGEMLHSQRLVTVERWKDFFQQDLAQLLPDLQIELEY
                     SPGFHSEVGLWQDLLNYHNKDVERRYTEYGPHRADLRLKTALGDADDVLSRGQKKLLM
                     MALKLSQIAMLHASNKETVVLLDDLTAELDSNAQRRLIERLSQLGSQVFITTLDHQAV
                     TQHLDGLSISYQLYNVDHGQVHAV"
     gene            4127. .6595
                     /gene="gyrB"
                     /locus_tag="ACIAD0004"
     CDS             4127. .6595
                     /gene="gyrB"
                     /locus_tag="ACIAD0004"
                     /EC_number="5.99.1.3"
                     /function="DNA bending, supercoiling, inversion
                     (MultiFun:3.1.1.1)"
                     /function="DNA replication (MultiFun:2.1.1)"
                     /function="transcription related (MultiFun:2.2.2)"
                     /note="Evidence 2a : Function of homologous gene

  Query Match             25.6%;  Score 170.8;  DB 1;  Length 110000;
  Best Local Similarity   54.7%;  Pred. No. 6.7e-24;
  Matches  362;  Conservative    0;  Mismatches  297;  Indels    3;  Gaps    1;

Qy          8 TGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAAC 67
              | ||||   ||| ||||||  |  | | |||| | |     || ||  ||   |    |
Db      90534 TTTATGCAGTTTATGGTGCCAGTGGATTTGGTCGTGGGGTTATGCCTCTTGCCAGAGCCC 90593

Qy         68 AAATAAAGCAAGAATGTGGTTCTGA---CTATGCTCTGGTTTTTGTGGATGATGTTTTGG 124
              |  | |||  | |   || | | ||    |||| | |||| ||| | |||||||
Db      90594 AGTTGAAGTCACAGCTTGATGCCGAAATGTATGATTTGGTCTTTATTGATGATGGCGAAC 90653

Qy        125 CAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTT 184
              | | | |  || ||||||  | | |||  |        |    |||||         |
Db      90654 GGGCAGATATCGTAAATGGTCATCAGGTGTATGATTTTGATCAATTTCTGGCTTTGGATG 90713

Qy        185 ATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTG 244
              |   |||| ||| | | | |||| ||||| |||  | |||| ||  | | | ||
Db      90714 CATCTGAAAAATATGTCAGTATTGCGATTGCCAATAGTCAGATTCGCGAAAAACTGATCA 90773

Qy        245 AGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTT 304
              |     ||  |   ||   | |||   ||   |  || |     || |       |
Db      90774 ACAAGCTACAACAGCATCAGATTGGCTCATGGAGCATCATTGCACCGACGGTGATCCAGA 90833

Qy        305 ATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTA 364
              |||  ||     ||| ||    || ||  |  ||| ||    |||     |  |  | |
Db      90834 TGGATGATGTCCAGATTGGAGAGGGTGCGGTGCTTTGTCATCATGTGCATTTGACCTCGA 90893

Qy        365 ATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTC 424
              |||  | ||| ||   |||||||||||| ||    |     || || ||||||||||||
Db      90894 ATATCCGTATTGGCCAGTTTTTTCATGCCAATTATTTTAGTTATGTGGCACATGATTGTG 90953

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              || || |||||||| || ||||| |||  ||   | || || ||  ||||   |||||
Db      90954 TGATTGGCGACTATGTGACCTTTGCACCTTCGGTACAGTGCAACGGCAATGTACATATTG 91013

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |||  |||| ||||| ||  | || ||| | |    ||||||   ||  | || || |
Db      91014 GCGACCATGCCTATATTGGGGCAGGGGCAATCTTGCGTCAGGGTACGCCAGACCGACCTC 91073

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTTCCTG 604
              |||| |||||||  || || ||||| || ||||| || || || ||   |  |||| |||
```

```
Db      91074 TTATCATTGGCGAAGGCGCGATTATCGGAATGGGTGCAGTGGTTACCCGAGATGTTGCTG 91133

Qy        605 CCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCTATTT 664
              |    |  ||||||   || ||||| || |||    |     |  ||| ||   |  |
Db      91134 CACACACGACTGTGGTGGGGAATCCTGCCAGAATTTTAACTGCCTATCCATCAGGAAGCT 91193

Qy        665 AA 666
              ||
Db      91194 AA 91195


RESULT 13
AL646082/c
LOCUS       AL646082              205050 bp    DNA     linear   BCT 16-APR-2005
DEFINITION  Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
            segment 7/11.
ACCESSION   AL646082 AL646053
VERSION     AL646082.1  GI:17431428
KEYWORDS    .
SOURCE      Ralstonia solanacearum
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
            Burkholderiaceae; Ralstonia.
REFERENCE   1
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Nature 415 (6871), 497-502 (2002)
   PUBMED   11823852
REFERENCE   2  (bases 1 to 205050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .205050
                     /organism="Ralstonia solanacearum"
                     /mol_type="genomic DNA"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
                     /plasmid="megaplasmid"
     gene            complement(486. .845)
                     /gene="RSp0956"
                     /note="synonym: RS02287"
     CDS             complement(486. .845)
                     /gene="RSp0956"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAD18107.1"
                     /db_xref="GI:17431429"
                     /db_xref="UniProt/TrEMBL:Q8XRA3"
                     /translation="MSEKLGAERLMPNEWRDDLEALDLEIVRFAIICQVKLLEPGVIQ
```

```
                RVLDNDARDCIHGNQTAFASLRGLLFLHFQMQKDLAEAYGPSDATEIVRRVWAHLQPR
                VGDQLGSLPNPGALSAV"
gene            1409. .1714
                /gene="RSp0957"
                /note="synonym: RS02288"
CDS             1409. .1714
                /gene="RSp0957"
                /function="miscellaneous; unknown"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="HYPOTHETICAL PROTEIN"
                /protein_id="CAD18108.1"
                /db_xref="GI:17431430"
                /db_xref="UniProt/TrEMBL:Q8XRA2"
                /translation="MEHDKYLHAGISAVIASGVTTLAVDSPNRFWYGLGAGMAVGLTK
                EIVDRRKTNGRFDSKDLLADLAGALIGAYAADSLLRPAVFKQPTGYAYGVQLNVAFD"
gene            complement(1869. .2540)
                /gene="RSp0958"
                /note="synonym: RS02289"
CDS             complement(1869. .2540)
                /gene="RSp0958"
                /function="miscellaneous; hypothetical/global homology"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="CONSERVED HYPOTHETICAL PROTEIN"
                /protein_id="CAD18109.1"
                /db_xref="GI:17431431"
                /db_xref="InterPro:IPR002063"
                /db_xref="InterPro:IPR007500"
                /db_xref="UniProt/TrEMBL:Q8XRA1"
                /translation="MPLLDQPLGHLARNIPGATGIFHEYQLDFCCGGQHSLRDAAQAK
                GIDAAPIAARLQALQAEATPDGAVDWSAVSPSMLIDHILERFHERHREQLPELIRLAR
                RVEHVHGDRPECPVGLSELLEAMWQELESHMQKEEQILFPMLARGHGLRAGGPIAVMR
                MEHDQHGEALQRLMTLTNDITPPRAACTTWRALYLGLSVFREDLMEHIHLENNILFEG
                AVAAD"
gene            2706. .4238
                /gene="RSp0959"
                /note="synonym: RS02290"
CDS             2706. .4238
                /gene="RSp0959"
                /function="miscellaneous; not classified regulator"
                /note="Product confidence : probable
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PROBABLE SIGMA-54 INTERACTING TRANSCRIPTION
                REGULATOR PROTEIN"
                /protein_id="CAD18110.1"
                /db_xref="GI:17431432"
                /db_xref="GOA:Q8XRA0"
                /db_xref="InterPro:IPR002078"
                /db_xref="InterPro:IPR002197"
                /db_xref="InterPro:IPR003018"
                /db_xref="InterPro:IPR003593"
                /db_xref="UniProt/TrEMBL:Q8XRA0"
                /translation="MIDRLLLADLVADLPPAVRMQRLVGSLRSHFRCGAVALLRLEED
                HLRPIAVDGLVRDALGRRFVVGQHPRLAAILARRGVTCFHHDSTLPDPYDGLIDEQAG
                APLPVHDCMGVSLHVEGQLWGALTLDTLKIGTFDADAQAELRQLAVVIEAAARTTRLE
                AEIRALRATRGDLPPGDPAADDWEIIGQSPAIHQLLHELEVVADSELPVLLLGETGVG
                KELFAHRLHRLSRRRDRPLVHVNCAALPESLAESELFGHVKGAFSGAVGERPGRFEAA
                DGGTLFLDEVGELPLPVQAKLLRALQNGEIQRLGADHPRHVDARVIAATNRNLRERVR
                DGAFRADLYHRLSVYPVPIPPLRERGNDVLLLAGRFLESNRARLGLRSLRLSPGAEEA
                LRRYPWPGNVRELEHVVSRAALKMLSRGASRDDIVTLETGLLDLDAQAVAPPSPAGLP
                PEASPDIAPGITLRDAVADCQRRHIQAALAIHQDNWAQTARTLGLDASNLHKLARRLG
                LK"
gene            complement(4338. .5294)
                /gene="RSp0960"
                /note="synonym: RS02291"
CDS             complement(4338. .5294)
                /gene="RSp0960"
                /EC_number="1.-.-.-"
                /function="miscellaneous; hypothetical/global homology"
                /note="Product confidence : putative
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PUTATIVE OXIDOREDUCTASE PROTEIN"
                /protein_id="CAD18111.1"
                /db_xref="GI:17431433"
                /db_xref="GOA:Q8XR99"
                /db_xref="UniProt/TrEMBL:Q8XR99"
                /translation="MGARVPKILVVGANGQIGSELALALAERYGHAQVITSDLVPAGR
                HPHLTHEMLNATDRGELAAVVERHGITQIYLLAAALSATGEKAPQWAWDLNMRSLLNV
                LEKARQMGVERVFWPSSIAVFGPTTPRDGTPQKTVMEPTTVYGISKQAGEGWCRWYHA
                NHGVDVRSMRYPGLISYKTLPGGGTTDYAVDIFHAAVKGEPYTCFLDPDEALPMMYMP
                DAIRATIELMEAPAERLTERGSYNIAGMTFTPAQIAAAIREHVPGFQIRYEPDYRQAI
                AQGWPDSIDDSVARADWGWQAEYGLESMVADMLVNLRTTLAA"
gene            complement(5294. .6493)
                /gene="kbl"
                /note="synonyms: RS02292, RSp0961"
CDS             complement(5294. .6493)
                /gene="kbl"
                /EC_number="2.3.1.29"
                /function="small molecule metabolism; central intermediary
                metabolism"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PROBABLE 2-AMINO-3-KETOBUTYRATE COENZYME A
                LIGASE PROTEIN"
                /protein_id="CAD18112.1"
                /db_xref="GI:17431434"
                /db_xref="GOA:Q8XR98"
                /db_xref="InterPro:IPR001917"
                /db_xref="InterPro:IPR003408"
                /db_xref="InterPro:IPR004839"
                /db_xref="InterPro:IPR011282"
                /db_xref="UniProt/TrEMBL:Q8XR98"
                /translation="MPNAEAFYAFIRTELDAIREAGLFKQERVIASPQGARVRTIDGR
                EVINLCANNYLGLSSHPRVIDAAHEALRTHGFGLSSVRFICGTQDLHKTLEARLSAFL
                GTEDTILYGSAFDANGGLFETLLGAEDAVISDALNHASIIDGIRLCKARRYRYQHNDL
                DDLRAQLTQARADGARHLLVFTDGVFSMDGTIARLDEMRSICDAHGALLGIDECHATG
                FLGPRGRGTHELRGVFGKVDIITGTLGKALGGASGGFTSARKEVVALLRQRSRPYLFS
                NTVAPAIIGASIAVLDILEGSTELRDRLERNTRFFRAGLEALGFDVKPGEHPIIPIMV
                YDADKAQQLAQRLLELGVYVVGFFYPVVPKGQARIRVQMSAVHDEAALQAALGAFGQA
                GHELGLI"

  Query Match          18.1%;  Score 120.6;  DB 1;  Length 205050;
  Best Local Similarity  51.6%;  Pred. No. 3.3e-14;
  Matches  276;  Conservative    0;  Mismatches  259;  Indels    0;  Gaps    0;

Qy      101 TGGTTTTTGTGGATGATGTTTTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAA 160
```

```
Db      62432 TCGTCTTCGTCGATGACGGCGCGGACGGCAGCGGCAGCAACGGGCACCGCGTGCTGACCT 62373

Qy        161 CCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATG 220

Db      62372 ATCCGCAGTTCCTGGCCGAGCCCGCCGCCGACAAGCGGCTGTGTTTTGCGATCGCCAACA 62313

Qy        221 ATAAGATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTA 280

Db      62312 GCCAGGTCCGCGAAAAGCTGGCAACGCGCGCCGTCGGCGACGGCATCGCGTGCATGGAGG 62253

Qy        281 TAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTT 340

Db      62252 TGCGCGGGGCCAACGCCGTCGTGCTGGACGCCGTGGAAATCGGCGCCGGCGCGGTGCTCT 62193

Qy        341 CTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATAT 400

Db      62192 GTCCGTTCGTGACGCTGACCAGCAATATCCGGATCGGCAAACACTTCCACGCCAACATCT 62133

Qy        401 ACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTA 460

Db      62132 ATGCGTACGTGGCGCACGATTGCGTGATCGGCGACTACGTCACGTTCGCGCCCGGGGCCA 62073

Qy        461 AATGTAATGGATATGTTGTTATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTA 520

Db      62072 AGTGCAACGGCAACGTCGTCATCGAGGACCACGCCTATGTCGGCACGGGCGCGGTGCTCA 62013

Qy        521 AGCAGGGTGTTCCTAATCGCCCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGG 580

Db      62012 AGCAGGGCAAGCCGGGCGCGCCGCTGGTCATCGGCAAGGGGGCCGTGGTCGGCATGGGCG 61953

Qy        581 CTGTTGTCACTAAAAGTGTTCCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAG 635

Db      61952 CGGTGGTGACCCGCGATGTGCCCGCCGGCGCGACGGTCGTCGGCAATCCCGCACG 61898


RESULT 14
CJ11168X5
LOCUS       CJ11168X5             263335 bp    DNA     linear   BCT 16-APR-2005
DEFINITION  Campylobacter jejuni subsp. jejuni NCTC 11168 complete genome;
            segment 5/6.
ACCESSION   AL139078 AL111168
VERSION     AL139078.2  GI:6968723
KEYWORDS    .
SOURCE      Campylobacter jejuni subsp. jejuni NCTC 11168
  ORGANISM  Campylobacter jejuni subsp. jejuni NCTC 11168
            Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
            Campylobacteraceae; Campylobacter.
REFERENCE   1
  AUTHORS   Parkhill,J., Wren,B.W., Mungall,K., Ketley,J.M., Churcher,C.,
            Basham,D., Chillingworth,T., Davies,R.M., Feltwell,T., Holroyd,S.,
            Jagels,K., Karlyshev,A.V., Moule,S., Pallen,M.J., Penn,C.W.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., van Vliet,A.H.,
            Whitehead,S. and Barrell,B.G.
  TITLE     The genome sequence of the food-borne pathogen Campylobacter jejuni
            reveals hypervariable sequences
  JOURNAL   Nature 403 (6770), 665-668 (2000)
   PUBMED   10688204
REFERENCE   2  (bases 1 to 263335)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2000) Submitted on behalf of the Campylobacter
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of C. jejuni sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/C_jejuni/).
FEATURES             Location/Qualifiers
     source          1. .263335
                     /organism="Campylobacter jejuni subsp. jejuni NCTC 11168"
                     /mol_type="genomic DNA"
                     /strain="NCTC 11168"
                     /db_xref="taxon:192222"
     gene            complement(38. .1369)
                     /gene="accC"
     CDS             complement(38. .1369)
                     /gene="accC"
                     /EC_number="6.4.1.2"
                     /EC_number="6.3.4.14"
                     /note="Cj1290c, accC, probable biotin carboxylase (subunit
                     of acetyl-CoA carboxylase (EC 6.4.1.2)), len: 443 aa;
                     highly similar to many e.g. ACCC_ECOLI biotin carboxylase
                     (EC 6.3.4.14) (449 aa), fasta scores; opt: 1560 z-score:
                     1741.0 E(): 0, 53.2% identity in 442 aa overlap. 67.0%
                     identity to HP0370. Also similar to Cj1037c pycA (46.5%
                     identity in 445 aa overlap). Contains PS00866 and PS00867
                     Carbamoyl-phosphate synthase subdomain signatures 1 and 2,
                     and Pfam match to entry PF00289 CPSase_L_chain,
                     Carbamoyl-phosphate synthase (CPSase)"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin carboxylase"
                     /protein_id="CAB73717.1"
                     /db_xref="GI:6968724"
                     /db_xref="GOA:Q9PN09"
                     /db_xref="InterPro:IPR004549"
                     /db_xref="InterPro:IPR005479"
                     /db_xref="InterPro:IPR005481"
                     /db_xref="InterPro:IPR005482"
                     /db_xref="InterPro:IPR011761"
                     /db_xref="InterPro:IPR011764"
                     /db_xref="UniProt/TrEMBL:Q9PN09"
                     /translation="MEIKSILIANRGEIALRALRTIKEMGKKAICVYSEADKDALYLK
                     YADASICIGKARSSESYLNIPAIIAAAEIAEADAIFPGYGFLSENQNFVEICAKHNIK
                     FIGPSVEAMNLMSDKSKAKQVMQRAGVPVIPGSDGALAGAEAAKKLAKEIGYPVILKA
                     AAGGGGRGMRVVENEKDLEKAYWSAESEAMTAFGDGTMYMEKYIQNPRHIEVQVIGDS
                     FGNVIHVGERDCSMQRRHQKLIEESPAILLDEKTRTRLHETAIKAAKAIGYEGAGTFE
                     FLVDKNLDFYFIEMNTRLQVEHCVSEMVSGIDIIEQMIKVAEGYALPSQESIKLNGHS
                     IECRITAEDSKTFLPSPGKITKYIPPAGRNVRMESHCYQDYSVPAYYDSMIGKLVVWA
                     EDRNKAIAKMKVALDELLISGIKTTKDFHLSMMENPDFINNNYDTNYLARH"
     misc_feature    complement(47. .1357)
                     /gene="accC"
                     /note="Pfam match to entry PF00289 CPSase_L_chain,
                     Carbamoyl-phosphate synthase (CPSase), score 640.70,
                     E-value 8e-189"
     misc_feature    complement(488. .511)
                     /gene="accC"
                     /note="PS00867 Carbamoyl-phosphate synthase subdomain
                     signature 2"
     misc_feature    complement(866. .910)
                     /gene="accC"
                     /note="PS00866 Carbamoyl-phosphate synthase subdomain
                     signature 1"
     gene            complement(1371. .1826)
                     /gene="accB"
     CDS             complement(1371. .1826)
                     /gene="accB"
                     /EC_number="6.4.1.2"
                     /note="Cj1291c, accB, probable biotin carboxyl carrier
                     protein of acetyl-CoA carboxylase, len: 151 aa; similar to
                     e.g. BCCP_ECOLI biotin carboxyl carrier protein of
                     acetyl-CoA carboxylase (EC 6.4.1.2) (156 aa), fasta
                     scores; opt: 307 z-score: 329.6 E(): 4.6e-11, 38.9%
                     identity in 162 aa overlap. 41.4% identity to HP0371.
                     Contains Pfam match to entry PF00364 biotin_req_enzy,
                     Biotin-requiring enzymes"
                     /codon_start=1
                     /transl_table=11
                     /product="putative biotin carboxyl carrier protein of
                     acetyl-CoA carboxylase"
                     /protein_id="CAB73718.1"
                     /db_xref="GI:6968725"
                     /db_xref="GOA:Q9PN08"
                     /db_xref="InterPro:IPR000089"
                     /db_xref="InterPro:IPR001249"
```

```
                     /db_xref="InterPro:IPR001638"
                     /db_xref="UniProt/TrEMBL:Q9PN08"
                     /translation="MTKEEIKELVNLFAEANISKIKIKEQDGFEIELERDMCCDVPAP
                     VCPPVPAPQPINVSVVNEAQPSQSVKSNKPSINSPMVGTFYQAPSPGAAPFVKVGSTV
                     KKGDTIAIIEAMKIMNEIEAEFDCRIAEILVADGQPVEFGMPLFAVEKL"
     misc_feature    complement(1383. .1592)
                     /gene="accB"
                     /note="Pfam match to entry PF00364 biotin_req_enzy,
                     Biotin-requiring enzymes, score 104.20, E-value 2.5e-27"
     gene            1974. .2534
                     /gene="dcd"
     CDS             1974. .2534
                     /gene="dcd"
                     /EC_number="3.5.4.13"
                     /note="Cj1292, dcd, possible deoxycytidine triphosphate
                     deaminase, len: 186 aa; similar to many predicted dcd
                     proteins, and to DCD_ECOLI deoxycytidine triphosphate
                     deaminase (EC 3.5.4.13) (193 aa), fasta scores; opt: 160
                     z-score: 203.5 E(): 0.00049, 28.8% identity in 160 aa
                     overlap. 71.4% identity to HP0372. Contains Pfam match to
                     entry PF00692 dUTPase"
                     /codon_start=1
                     /transl_table=11
                     /product="possible deoxycytidine triphosphate deaminase"
                     /protein_id="CAB73719.1"
                     /db_xref="GI:6968726"
                     /db_xref="GOA:Q9PN07"
                     /db_xref="InterPro:IPR003232"
                     /db_xref="InterPro:IPR008180"
                     /db_xref="UniProt/Swiss-Prot:Q9PN07"
                     /translation="MGLKADNWIRKMALEHKMIEPFCEANIGKGVVSYGLSSYGYDIR
                     VGREFKIFTNVNSTVVDPKNFVEENVVDFEGDVCIVPANSFALARTIEYFKMPDNVLA
                     ICLGKSTYARCGIIVNVTPFEPGFEGHITIEISNTTPLPAKIYANEGIAQVLFLQGDE
                     KCDTTYKDKKGKYQAQTGITLPRILK"
     misc_feature    2046. .2522
                     /gene="dcd"
                     /note="Pfam match to entry PF00692 dUTPase, dUTPase, score
                     3.90, E-value 0.0016"
     gene            2584. .3588
                     /gene="Cj1293"
     CDS             2584. .3588
                     /gene="Cj1293"
                     /note="Cj1293, possible sugar nucleotide
                     epimerase/dehydratase, len: 334 aa; similar to many e. .g
                     TR:Q45984 (EMBL:U27301) Caulobacter crescentus FLAA1
                     protein (331 aa), fasta scores; opt: 1240 z-score: 1410.9
                     E(): 0, 57.8% identity in 329 aa overlap, CAPD_STAAU CAPD
                     protein (599 aa), fasta scores; opt: 529 z-score: 602.6
                     E(): 2.9e-26, 33.9% identity in 319 aa overlap, and
                     TR:O69130 (EMBL:AF064070) Burkholderia pseudomallei
                     putative epimerase/dehydratase WBII (637 aa), fasta
                     scores; opt: 433 z-score: 493.6 E(): 3.4e-20, 30.5%
                     identity in 282 aa overlap. 64.6% identity to HP0840.
                     Contains PS00017 ATP/GTP-binding site motif A (P-loop"
                     /codon_start=1
                     /transl_table=11
                     /product="possible sugar nucleotide epimerase/dehydratase"
                     /protein_id="CAB73720.1"
                     /db_xref="GI:6968727"
                     /db_xref="GOA:Q9PN06"
                     /db_xref="InterPro:IPR003869"
                     /db_xref="UniProt/TrEMBL:Q9PN06"
                     /translation="MFNKKNILITGGTGSFGKTYTKVLLENYKPNKIIIYSRDELKQF
                     EMASVFNAPCMRYFIGDVRDKERLSAAMRDVDFVIHAAAMKHVPIAEYNPMECIKTNI
                     HGAQNVIDACFENGVKKCIALSTDKACNPVNLYGATKLASDKLFVAANNIAGNKQTRF
                     GVTRYGNVVGSRGSVVPFFKKLISEGAKELPITDTRMTRFWISLEDGVKFVLSNFERM
                     HGGEIFIPKIPSMKITDLAHALAPNLSHKIIGIRAGEKLHEIMISSDDSHLTYEFENY
                     YAISPSIKFVDKDNDFSINALGEKGQKVKDGFSYSSDNNPLWASEKELLEIINHTEGF
                     "
     misc_feature    2617. .2640
                     /gene="Cj1293"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     gene            3590. .4720
                     /gene="Cj1294"
     CDS             3590. .4720
                     /gene="Cj1294"
                     /note="Cj1294, probable aminotransferase, len: 376 aa;
                     similar to many members of the degT family e.g. SPSC_BACSU
                     spore coat polysaccharide biosynthesis protein (389 aa),
                     fasta scores; opt: 586 z-score: 676.7 E(): 2.1e-30, 35.3%
                     identity in 385 aa overlap, and TR:O88001 (EMBL:AJ007747)
                     Bordetella bronchiseptica putative amino-sugar
                     biosynthesis protein WLBC (366 aa), fasta scores; opt: 482
                     z-score: 558.1 E(): 8.7e-24, 28.8% identity in 330 aa
                     overlap. 44.8% identity to HP0366 . Contains Pfam match to
                     entry PF01041 DegT_DnrJ_EryC1, DegT/DnrJ/EryC1/StrS
                     family"
                     /codon_start=1
                     /transl_table=11
                     /product="putative aminotransferase (degT family)"
                     /protein_id="CAB73721.1"
                     /db_xref="GI:6968728"

  Query Match          12.8%;  Score 85.2;  DB 1;  Length 263335;
  Best Local Similarity  54.5%;  Pred. No. 2.4e-07;
  Matches  171;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy       189 AAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTC 248
             ||||  | ||||  | | | | | |  |    ||  || || | | | |  ||
Db     28754 AAAAGCATCTTTTGCTTTAGGTGTAGGACAAATCAATTTAAGAAAAAAAATCCTTGTAAA 28813

Qy       249 AATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGA 308
             | | | |  | |       | |    || ||| | | ||||||  |      ||||||  |
Db     28814 ACTTTCAAAAAAATCATGTAATTTTATCACTTTTATACATCCTCAATCTTTTGTTTCAAA 28873

Qy       309 TCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATAC 368
               |  |||  |||||    || | |||  | | ||| ||||| || ||   | ||||| |
Db     28874 AGAAGCTAAAATAGGACAAGGGGTTATAGTGTGTCCTTTTGTAACCATTAATGCTAATTC 28933

Qy       369 TCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAAT 428
               |||||||    ||| | | |   || || || |||  | | || |||||||||| || |
Db     28934 AAATATAGGAGATTTTGTGCTTTGTAATATTTATTCAAGCATAGCTCATGATTGTAAAGT 28993

Qy       429 AGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGA 488
             ||||||      ||  ||    |||   |||| |  ||||||| ||      || | | |
Db     28994 AGGAGAAGGAAGTATTTTAAGCCCTTATGCTACACTTAATGGAAATTCAAGCATAGGAAA 29053

Qy       489 CAATGCATATATAG 502
               |||   | | |||
Db     29054 AAATTGTTTTTTAG 29067


RESULT 15
AE013035
LOCUS       AE013035               10722 bp    DNA     linear   BCT 10-JUN-2004
DEFINITION  Thermoanaerobacter tengcongensis MB4, section 62 of 244 of the
            complete genome.
ACCESSION   AE013035 AE008691
VERSION     AE013035.1  GI:20515651
KEYWORDS    .
SOURCE      Thermoanaerobacter tengcongensis MB4
  ORGANISM  Thermoanaerobacter tengcongensis MB4
            Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
            Thermoanaerobacteriaceae; Thermoanaerobacter.
REFERENCE   1  (bases 1 to 10722)
  AUTHORS   Bao,Q., Tian,Y., Li,W., Xu,Z., Xuan,Z., Hu,S., Dong,W., Yang,J.,
            Chen,Y., Xue,Y., Xu,Y., Lai,X., Huang,L., Dong,X., Ma,Y., Ling,L.,
            Tan,H., Chen,R., Wang,J., Yu,J. and Yang,H.
  TITLE     A Complete Sequence of the T. tengcongensis Genome
  JOURNAL   Genome Res. 12 (5), 689-700 (2002)
   PUBMED   11997336
REFERENCE   2  (bases 1 to 10722)
  AUTHORS   Bao,Q., Xu,Z., Hu,S., Dong,W., Chen,Y., Wang,J., Yu,J. and Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and
            Bioinformatics Center, Institute of Genetics and Development,
```

```
            Chinese Academy of Sciences, Beijing Airport Industrial Zone B6,
            Beijing 101300, China
REFERENCE   3  (bases 1 to 10722)
  AUTHORS   Li,W., Xuan,Z., Yang,J., Ling,L. and Chen,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of
            Sciences, Beijing 100101, China
REFERENCE   4  (bases 1 to 10722)
  AUTHORS   Tian,Y., Xue,Y., Xu,Y., Lai,X., Huang,L., Dong,X., Ma,Y. and Tan,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy
            of Sciences, Beijing 100080, China
FEATURES             Location/Qualifiers
     source          1. .10722
                     /organism="Thermoanaerobacter tengcongensis MB4"
                     /mol_type="genomic DNA"
                     /strain="MB4"
                     /db_xref="taxon:273068"
     gene            133. .761
                     /locus_tag="TTE0663"
     RBS             133. .137
                     /locus_tag="TTE0663"
     CDS             144. .761
                     /locus_tag="TTE0663"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAM23929.1"
                     /db_xref="GI:20515652"
                     /translation="MNNFAEELTYWYFRLNGFFLLQNYVLHNIGQGEQRGTADCDLLA
                     IRFPYVYEKVGGQNKDWDKEKFSTWGFDIDRDNLALIIEVKSGRELKKNKIEKSFSPE
                     RVKTALYRFGIFSHKDVHRIHENLLKQKYVKESSWTIAKVLVTEKAKEIEGNWLNLSL
                     DEIDKFIHKRIKSYLHDKYADRMYFPSILIQYIAWKDKNRILNDC"
     gene            791. .1605
                     /locus_tag="TTE0664"
     RBS             791. .795
                     /locus_tag="TTE0664"
     CDS             805. .1605
                     /locus_tag="TTE0664"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAM23930.1"
                     /db_xref="GI:20515653"
                     /translation="MARKFSDRTMAICGEILSELFSDGELQNFLSRFGLLDIYRRLPQ
                     GTGKLKRIQEVIAYLNQLNNTENKDKIWDDIVIETVKTIKQIEERAPVSSLVMEKIRK
                     LEECLALDGYQIENGKIIKALPAPTETIHIRDRLEQKLLDLGWFVPLQHLQDARENYA
                     EERWAAANASIRSFLQAIFDNIAEAMPNFPKEELVPGGARRQFLETIGFLTKEEAELV
                     RALFKILATKGSHPGLSDESDCSSRILMSIGIAWRFLSRFLEKGGNFS"
     gene            1587. .2207
                     /gene="WcaJ"
                     /locus_tag="TTE0665"
     RBS             1587. .1591
                     /gene="WcaJ"
                     /locus_tag="TTE0665"
     CDS             1602. .2207
                     /gene="WcaJ"
                     /locus_tag="TTE0665"
                     /note="Best Blastp hit =  gi|7449885|pir||D70037 capsular
                     polysaccharide biosynthesis homolog yvfC - Bacillus
                     subtilis gi|1495291|emb|CAA96480.1| (Z71928) hypothetical
                     protein [Bacillus subtilis] gi|1945702|emb|CAB07996.1|
                     (Z94043) hypothetical protein [Bacillus subtilis]
                     gi|2635938|emb|CAB15430.1| (Z99121) similar to capsular
                     polysaccharide biosynthesis [Bacillus subtilis], score
                     243, E-value 2.00E-63"
                     /codon_start=1
                     /transl_table=11
                     /product="Sugar transferases involved in
                     lipopolysaccharide synthesis"
                     /protein_id="AAM23931.1"
                     /db_xref="GI:20515654"
                     /translation="MKLIIKRLIDIFVSLFLLIILSPFLVIISLIVFVAMGPPVIYKQ
                     TRAGFKGKPFTIYKFRTMTNEKDEYGNLLPDEKRLTKMGKFLRSTSLDELPELFNVLK
                     GDMSLVGPRPLLMEYLNYYTEEQMRRHDVKPGITGWAQVNGRNSLSWEEKFKLDVWYV
                     DNWSLWLDFKILFLTLIKVIKREGLSAEGYATMPKFTGSKD"
     misc_feature    1617. .2159
                     /gene="WcaJ"
                     /locus_tag="TTE0665"
                     /note="Pfam match to entry Bact_transf, Bacterial sugar
                     transferase, score 209.4, E-value 5.60E-59"
     gene            2215. .2888
                     /gene="WbbJ2"
                     /locus_tag="TTE0666"
     RBS             2215. .2219
                     /gene="WbbJ2"
                     /locus_tag="TTE0666"
     CDS             2229. .2888
                     /gene="WbbJ2"
                     /locus_tag="TTE0666"
                     /note="Best Blastp hit =  gi|7475816|pir||E70037 serine
                     O-acetyltransferase homolog yvfD - Bacillus subtilis
                     gi|1495292|emb|CAA96481.1| (Z71928) hypothetical protein
                     [Bacillus subtilis] gi|1945703|emb|CAB07997.1| (Z94043)
                     hypothetical protein [Bacillus subtilis]
                     gi|2635937|emb|CAB15429.1| (Z99121) similar to serine
                     O-acetyltransferase [Bacillus subtilis], score 73.2,
                     E-value 3.00E-12"
                     /codon_start=1
                     /transl_table=11
                     /product="Acetyltransferases (the isoleucine patch
                     superfamily)"
                     /protein_id="AAM23932.1"
                     /db_xref="GI:20515655"
                     /translation="MLKDLFIIGAGGVGKEVAWIVEEINNINPIWNIVGFIDDNKDLH
                     DRTINGVKVLGGIEYLLNLKRKIYAVCAIADYKSRKRIVTKLERYDHIEFVNIIHPSV
                     YISKTNHLGKGVIIYPGCILTTNIKIGDHVIISPKCGIGHETFIDDFVTLLWDVNVAG
                     NVVINEGCLIGSGATIIQNKQIGKGVIIGAGAVVVDDIPPYCTAVGVPAKPIKFHEEA
                     K"
     gene            2875. .4007
                     /gene="WecE2"
                     /locus_tag="TTE0667"
     RBS             2875. .2879
                     /gene="WecE2"
                     /locus_tag="TTE0667"
     CDS             2889. .4007
                     /gene="WecE2"
                     /locus_tag="TTE0667"
                     /note="Best Blastp hit =  gi|7433469|pir||G72359
                     hypothetical protein TM0572 - Thermotoga maritima (strain
                     MSB8) gi|4981089|gb|AAD35657.1|AE001732_2 '(AE001732)
                     lipopolysaccharide biosynthesis protein, putative
                     [Thermotoga maritima)', score 329, E-value 4.00E-89"
                     /codon_start=1
                     /transl_table=11
                     /product="predicted pyridoxal phosphate-dependent enzyme
                     apparently involved in regulation of cell wall biogenesis"
                     /protein_id="AAM23933.1"
                     /db_xref="GI:20515656"
                     /translation="MKVPLSRPDITQKEIDAVVEVLRSDILSIGPKIEEFERKIADYV
                     GKKYAIAVNSGTSALHLIVRALGIKDGDEVITTPFTFIASVNCFLFERAKPVFVDIDP
                     DTLNMDINKIEEKITEKTKAILTVDVFGQPMDMEKVNEIAKKYGLKVIEDSCEALGSE
                     YKGIKAGTLCDAGAFAFYPNKQITTAEGGVIVTDDEEIAQLCRSMRSQGRPITGLWLE
                     HERLGFNYRLSELHAALGIVQLERIEEIIKRRLEIAEKYNERLKDVKGVKIPYIAPEV
                     NKMSWFVYVIRVDENIDRDKVMQYLIDNGIGCRPYFTPVHLQSYIKELTGHKEGDFPI
                     TEKVARSTIALPFFNKITDEQIDYVVDKLKEAIKIYSR"
     misc_feature    2910. .3980
                     /gene="WecE2"
                     /locus_tag="TTE0667"
                     /note="Pfam match to entry DegT_DnrJ_EryC1,
                     DegT/DnrJ/EryC1/StrS family, score 548.7, E-value
                     4.00E-161"
     gene            4010. .5894
                     /locus_tag="TTE0668"
     RBS             4010. .4014
                     /locus_tag="TTE0668"
```

```
     CDS             4023. .5894
                     /locus_tag="TTE0668"
                     /note="Best Blastp hit =  gi|10176342|dbj|BAB07437.1|
                     (AP001519) capsular polysaccharide biosynthesis [Bacillus
                     halodurans], score 482, E-value 1.00E-135"
                     /codon_start=1
                     /transl_table=11
                     /product="predicted nucleoside-diphosphate sugar
                     epimerases"
                     /protein_id="AAM23934.1"
                     /db_xref="GI:20515657"
                     /translation="MWKNKRKILLLLIDIILLNIAYFLAFYLRFDYNIPHQYFLLYKK
                     TFYIVVTVTVVTFIFFKQYNKIWRYASLHDIMDLIIFTTIGSAMAAFVLLFLQIHLPR
                     SIYPLYWFLSALLIGGSRIAYRSYMENSSKVKVKNIKYKKVLIVGAGDAGRILVREIK
                     NHPETGLKPVAIVDDDVSKLHRSIDDVPVVGKIDDIQEVITEKEIDEVIFAIPSAPRT
                     VLKKVIDKCSEMRIKVKTLPAIYELVDGTISISSIRDVDINDLLGREPVKIDLEEISS
                     YIKGKKVMVTGGGGSIGSELCRQIARFSPSRLIILDIYENNMYEVEQELKRKYPELDV
                     ICLVANIREEKKINKIFSEYTPDIVFHAAAHKHVPLMEVSPHEAIKNNVYGTLNLVKV
                     SDLYNVNRFVMISTDKAVNPTSVMGASKRICEMIIQEYDKKSKTEFVAVRFGNVLGSN
                     GSVIPLFKKQIAEGGPVTVTHEEIKRYFMTIPEAVQLVIQAGAMAKGGEIFVLDMGEP
                     VKIIDLARTLIRLSGFEPDVDIPIKIIGLRPGEKLFEELLISDDKYVSTKHEKIFIEK
                     PVSVDYGDLILTLEKYKDKIDDMDEDEIREFIKSLVPEYNPDYQRSLMEEVAATEIE"
     misc_feature    4257. .5834
                     /locus_tag="TTE0668"
                     /note="Pfam match to entry Polysacc_synt_2, Polysaccharide
                     biosynthesis protein, score 817.5, E-value 4.90E-242"
     misc_feature    4869. .5699
                     /locus_tag="TTE0668"
                     /note="Pfam match to entry Epimerase, NAD dependent
                     epimerase/dehydratase f, score -62.2, E-value 1.40E-08"
     gene            5929. .6354
                     /locus_tag="TTE0669"
```

Query Match 9.9%; Score 65.8; DB 1; Length 10722;
Best Local Similarity 46.9%; Pred. No. 0.0036;
Matches 276; Conservative 0; Mismatches 307; Indels 6; Gaps 2;

```
Qy      15 TATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAATAAA 74
           |||| |||| || ||  |    |||| |||           ||| || |  ||  ||  ||
Db    2246 TATTATTGGAGCAGGTGGGGTAGGTAAAGAGGTTGCGTGGATTGTAGAAGAAATAAATAA 2305

Qy      75 GCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGAT---GTTTTGGCAGGAAA 131
             |  |   ||   |   || |    || |||  ||| ||   | ||||   |  |
Db    2306 TATAAATCCTATTTGGAATATTGTTGGCTTTATTGACGATAATAAAGATTTGCATGATAG 2365

Qy     132 GAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAA 191
           |   | || ||  |  |||||||  |   |    | | ||  |   | | |    ||
Db    2366 AACTATAAACGGGGTAAAAGTGCTAGGAGGAATTGAGTATTTATTAAATCTTAAGAGGAA 2425

Qy     192 AAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGA---GTGTCTGAGTC 248
           |  |||  |     ||| |||||  || ||||     || | |||   ||  |  |
Db    2426 GATTTATGCTGTATGTGCAATTGCAGATTATAAATCTAGAAAAAGAATTGTTACAAAACT 2485

Qy     249 AATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGA 308
           |  |   |  ||    | |||    |||  ||||  ||||||  |  |  | | |    |
Db    2486 AGAACGTTATGACCATATAGAATTTGTAAACATAATTCATCCATCTGTCTATATCTCAAA 2545

Qy     309 TCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATAC 368
           | |    | ||||  || |  || || | |||   |   | | || | || ||||
Db    2546 AACAAATCATCTTGGTAAGGGGGTAATAATATATCCAGGTTGCATATTAACCACGAATAT 2605

Qy     369 TCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAAT 428
           | ||||||    ||  ||       | |   | ||| ||||||   |   ||
Db    2606 AAAAATAGGGGACCATGTTATAATAAGTCCCAAATGTGGCATTGGACATGAAACTTTTAT 2665

Qy     429 AGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGA 488
           |  || | ||||||  |   |   | |  ||   |   ||| |||| || ||  | ||
Db    2666 TGATGATTTTGTTACTCTACTTTGGGATGTAAATGTTGCAGGAAATGTAGTAATAAATGA 2725

Qy     489 CAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTAT 548
           |   |  ||||| || |||||    || |  ||       | | | |   | | |
Db    2726 AGGTTGTTTAATAGGTTCAGGTGCTACGATAATACAAAACAAACAAATTGGAAAAGGTGT 2785
```

```
Qy     549 TATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGT 597
           |||   |  | ||      ||| | | |  | | |     || |  ||
Db    2786 AATTATAGGTGCAGGAGCGGTAGTTGTTGATGATATACCACCTTATTGT 2834
```

Search completed: May 10, 2006, 20:33:53
Job time : 2423.44 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:42:47 ; Search time 310.898 Seconds
(without alignments)
14276.972 Million cell updates/sec

Title: US-09-423-093B-2_COPY_13156_13821
Perfect score: 666
Sequence: 1 atgaatttgtatggtatttt.....gatcgccaacatctatttaa 666

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	666	100.0	14024	2	AAX06749	Aax06749 E. coli O
2	666	100.0	14024	3	AAZ56386	Aaz56386 Escherich
c 3	666	100.0	14181	10	ADC01214	Adc01214 Enterohae
c 4	666	100.0	14187	9	ACD19108	Acd19108 E. coli 0
5	218.2	32.8	687	9	ADA29465	Ada29465 DNA encod
6	194.4	29.2	14256	14	AEC31451	Aec31451 E. coli t
7	71.4	10.7	1199	8	ACA47793	Aca47793 Prokaryot
c 8	47.6	7.1	110000	14	AEB42401_32	Continuation (33 o
9	46.2	6.9	624	12	ADH88991	Adh88991 Escherich
10	45.2	6.8	645	5	AAH65371	Aah65371 C glutami
11	45.2	6.8	645	8	ACA02087	Aca02087 C. glutam
12	45.2	6.8	349980	5	AAH68525	Aah68525 C glutami
13	44.6	6.7	609	8	ACF74035	Acf74035 Staphyloc
14	44.6	6.7	627	4	AAS51946	Aas51946 Staphyloc
15	44.6	6.7	642	8	ACA20258	Aca20258 Prokaryot
16	44.6	6.7	681	4	AAS54720	Aas54720 Staphyloc
17	44.6	6.7	681	4	AAS55048	Aas55048 Staphyloc
18	44.6	6.7	3700	14	ADW94081	Adw94081 Staphyloc
19	44.6	6.7	4350	14	ADW94080	Adw94080 Staphyloc
c 20	44.6	6.7	6876	2	AAV74497	Aav74497 Staphyloc
21	44.4	6.7	657	10	ADE99826	Ade99826 Bacterial
22	44.2	6.6	15958	14	AEB15068	Aeb15068 E. coli O
23	43.4	6.5	4605	10	ADF00725	Adf00725 Bacterial
24	43.2	6.5	110000	14	AEB42401_00	Aeb42401 L. pneumo
c 25	43	6.5	720	4	AAH53151	Aah53151 S. epider
c 26	43	6.5	972	6	ABN92900	Abn92900 Staphyloc
c 27	43	6.5	972	13	ADS01165	Ads01165 Staphyloc
28	43	6.5	3572	4	AAH54624	Aah54624 S. epider
29	43	6.5	3806	4	AAH54802	Aah54802 S. epider
30	42	6.3	720	8	ACA47130	Aca47130 Prokaryot
31	41.8	6.3	2013	6	ABL56229	Abl56229 AmEPV VET
c 32	41.8	6.3	50000	6	ABL56202	Abl56202 AmEPV gen
33	41.8	6.3	110000	6	ABQ69245_10	Continuation (11 o
34	41.8	6.3	110000	6	ABQ67197_09	Continuation (10 o
35	41.6	6.2	660	8	ACA37124	Aca37124 Prokaryot
36	41.6	6.2	1485	10	ADF00316	Adf00316 Bacterial
c 37	41.6	6.2	1500	14	ADZ27256	Adz27256 Barrel me
38	41.6	6.2	130312	14	AEB39168	Aeb39168 L. pneumo
c 39	41.6	6.2	143354	14	AEB42740	Aeb42740 L. pneumo
c 40	41.4	6.2	3422	6	AAL48669	Aal48669 Human apo
41	41.2	6.2	822	9	ADA31862	Ada31862 DNA encod
42	41	6.2	6763	6	ABQ67068	Abq67068 Human ang
43	41	6.2	29607	4	AAL37261	Aal37261 Human mus
44	41	6.2	29607	8	ABX60249	Abx60249 cDNA enco
45	41	6.2	29607	12	ADJ30999	Adj30999 Human mus

ALIGNMENTS

RESULT 1
AAX06749
ID AAX06749 standard; DNA; 14024 BP.
XX
AC AAX06749;
XX
DT 26-APR-1999 (first entry)
XX
DE E. coli O157 antigen gene cluster.
XX
KW O antigen; O157 antigen; diagnosis; wbdN gene; wzx gene; wzy gene;
KW wbdP gene; wbdO gene; wbdR gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 79. .861
FT /*tag= a
FT /product= "sugar transferase"
FT /note= "wbdN gene (ORF1), this region is specifically
FT claimed in Claim 6; encodes AAW88312"
FT CDS 858. .2042
FT /*tag= b
FT /product= "O_antigen-polymerase"
FT /note= "wzy gene (ORF2), this region is specifically
FT claimed in Claim 6; encodes AAW88313"
FT CDS 2011. .2757
FT /*tag= c
FT /product= "sugar_transferase"
FT /note= "wbdO gene (ORF3), this region is specifically
FT claimed in Claim 6; encodes AAW88314"
FT CDS 2744. .4135
FT /*tag= d
FT /product= "O_antigen_flippase"
FT /note= "wzx gene (ORF4), this region is specifically
FT claimed in Claim 6; encodes AAW88315"
FT CDS 4132. .5232
FT /*tag= e
FT /product= "perosamine_synthetase"
FT /note= "per gene (ORF5); encodes AAW88316"
FT CDS 5257. .6471

```
FT                   /*tag=  f
FT                   /product= "sugar_transferase"
FT                   /note= "wbdP gene (ORF6), this region is specifically
FT                   claimed in Claim 6; encodes AAW88317"
FT   CDS             6491. .7609
FT                   /*tag=  g
FT                   /note= "ORF7; encodes AAW88318"
FT   CDS             7606. .8578
FT                   /*tag=  h
FT                   /note= "ORF8; encodes AAW88319"
FT   CDS             8580. .9089
FT                   /*tag=  i
FT                   /note= "ORF9; encodes AAW88320"
FT   CDS             9071. .10519
FT                   /*tag=  j
FT                   /note= "manC gene (ORF10); encodes AAW88321"
FT   CDS             10523. .11893
FT                   /*tag=  k
FT                   /note= "manD gene (ORF11); encodes AAW88322"
FT   repeat_region   12007. .13123
FT                   /tag= l
FT                   /note= "H-repeat, may play a role in gene cluster
FT                   assembly"
FT   CDS             13156. .13821
FT                   /*tag=  m
FT                   /product= "N-acetyl_transferase"
FT                   /note= "wbdR gene (ORF12), this region is specifically
FT                   claimed in Claim 6; encodes AAW88323"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88312, AAW88313, AAW88314, AAW88315, AAW88316, AAW88317,
DR   AAW88318, AAW88319, AAW88320, AAW88321, AAW88322, AAW88323.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 6; Fig 8; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Escherichia coli O157 O antigen. The gene cluster was
CC   obtained by PCR amplification of E. coli O157:H7 strain (C664-1992) DNA
CC   using primers (see AAX06752-63) based on O antigen gene sequences. 12
CC   Open reading frames within the gene cluster were identified on the basis
CC   of homology to known sequences. The genes encode proteins (see AAW88312-
CC   23) involved the synthesis of sugars present in the polysaccharide
CC   antigen, and in the transport or processing of polysaccharide or
CC   oligosaccharide units. The use of nucleic acid molecules derived from
CC   particular assembly and transport genes, particularly wbd (transferase),
CC   wzx (flippase) and wzy (polymerase) genes, within O antigen gene clusters
CC   improves the specificity of methods for the detection and identification
CC   of O antigens, e.g. in tests of food- or faecal- derived samples, or
CC   samples from patients. The O antigen is a major virulence factor of
CC   enteropathogenic E. coli strains that cause diarrhoea and haemorrhagic
CC   colitis
XX
SQ   Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 666;  DB 2;  Length 14024;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-165;

  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13156 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 13215

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13216 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 13275

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13276 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 13335

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13336 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 13395

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13396 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 13455

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13456 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 13515

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13516 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 13575

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13576 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 13635

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13636 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 13695

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13696 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 13755

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13756 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 13815

Qy        661 ATTTAA 666
              ||||||
Db      13816 ATTTAA 13821


RESULT 2
AAZ56386
ID   AAZ56386 standard; DNA; 14024 BP.
XX
AC   AAZ56386;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:56.
XX
KW   Flagellin; fliC; antigen; detection; ds.
XX
OS   Escherichia coli.
XX
PN   WO9961458-A1.
XX
PD   02-DEC-1999.
XX
PF   21-MAY-1999;   99WO-AU000385.
XX
PR   21-MAY-1998;   98AU-00003634.
XX
```

```
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 2000-072598/06.
XX
PT   Novel nucleic acid molecule useful for the detection of flagellated
PT   bacterial strains in food, feces, etc.
XX
PS   Claim 3; Page 229-233; 245pp; English.
XX
CC   AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all or
CC   part of an Escherichia coli flagellin protein except a protein expressed
CC   by E. coli H1, H7, H12 or H48 type strains. The present invention also
CC   describes a method of detecting the presence of E. coli of a particular H
CC   serotype in a sample, comprising specifically hybridising a nucleic acid,
CC   preferably at least a pair, derived from a flagellating gene, specific
CC   for a particular flagellin gene associated with the H serotype, to any
CC   E.coli in the sample which contain the gene, and detecting any hybridised
CC   molecules, identifying the presence of that serotype in the sample. (I)
CC   are useful for: (1) detecting the presence of E. coli of H serotype in a
CC   sample by hybridising at least one or a pair of (I) to any E. coli in the
CC   sample and detecting the hybridised nucleic acid molecules; and (2) for
CC   detecting the presence of both O and H-serotypes of E. coli by
CC   hybridising at least one or a pair of (I) to any E. coli present in the
CC   sample and detecting the hybridised nucleic acid molecules. (I) is
CC   particularly useful for detecting the combination of O and H antigen.
CC   Hybridised (I) when using at least one (I) is detected by southern blot
CC   analysis and, when using a pair of (I), is detected by polymerase chain
CC   reaction (PCR). AAZ56399 to AAZ56420 represent primers used in the
CC   exemplification of the present invention
XX
SQ   Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 666;  DB 3;  Length 14024;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-165;
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13156 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 13215

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13216 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 13275

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13276 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 13335

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13336 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 13395

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13396 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 13455

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13456 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 13515

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13516 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 13575

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13576 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 13635

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13636 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 13695

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13696 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 13755

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13756 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 13815

Qy        661 ATTTAA 666
              ||||||
Db      13816 ATTTAA 13821


RESULT 3
ADC01214/c
ID   ADC01214 standard; DNA; 14181 BP.
XX
AC   ADC01214;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Enterohaemorragic E. coli 0157:H7-specific nucleic acid SEQ ID NO: 1258.
XX
KW   ds; gene; enterohaemorragic; anti-bacterial.
XX
OS   Escherichia coli; 0157:H7.
XX
PN   JP2002355074-A.
XX
PD   10-DEC-2002.
XX
PF   24-JAN-2002; 2002JP-00015959.
XX
PR   24-JAN-2001; 2001JP-00112010.
XX
PA   (UYTS-) UNIV TSUKUBA.
XX
DR   WPI; 2003-451640/43.
XX
PT   Enterohemorragic Escherichia coli O157:H7-specific nucleic acid molecule
PT   and a polypeptide and its use, a polypeptide, a vector and a host cell.
XX
PS   Claim 2; SEQ ID NO 1258; 2067pp; Japanese.
XX
CC   The invention relates to a novel enteroHaemorragic Escherichia coli
CC   O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
CC   has anti-bacterial activity. The polypeptide can be used in detection
CC   and/or treatment of O157:H7 infection. The nucleotide sequence of the
CC   genome of Enterohaemorragic E coli O157:H7 was determined. The present
CC   sequence represents an E. coli 0157:H7-specific nucleic acid of the
CC   invention.
XX
SQ   Sequence 14181 BP; 4450 A; 2856 C; 2281 G; 4594 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 666;  DB 10;  Length 14181;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-165;
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 781

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 721

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 661

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        660 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 601

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 541

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 481

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 421

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 361

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 301

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 241

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 181

Qy        661 ATTTAA 666
              ||||||
Db        180 ATTTAA 175
```

RESULT 4
ACD19108/c
ID ACD19108 standard; DNA; 14187 BP.
XX
AC ACD19108;
XX
DT 27-OCT-2003 (revised)
DT 21-AUG-2003 (first entry)
XX
DE E. coli 0157 unique DNA sequence OZID_121.
XX
KW OZID; ds; acute haemorrhagic colitis; haemolytic uraemic syndrome;
KW food poisoning.
XX
OS Escherichia coli; strain 0157:H7.
XX
PN US2003023075-A1.
XX
PD 30-JAN-2003.
XX
PF 01-APR-2002; 2002US-00114170.
XX
PR 04-DEC-1998; 98US-0110955P.
PR 03-DEC-1999; 99US-00453702.
XX
PA (BLAT/) BLATTNER F R.
PA (BURL/) BURLAND V D.
PA (PERN/) PERNA N T.
PA (PLUN/) PLUNKETT G.
PA (WELC/) WELCH R.
XX
PI Blattner FR, Burland VD, Perna NT, Plunkett G, Welch R;
XX
DR WPI; 2003-479497/45.
XX
PT New DNA sequences from Escherichia coli strain O157:H7, useful for
PT detecting E. coli O157:H7 in a sample, or in designing diagnostic probes
PT which can be used to distinguish strain O157:H7 from strain K12 using
PT molecular techniques.
XX
PS Claim 16; SEQ ID NO 121; 33pp; English.
XX
CC The invention relates to an isolated DNA molecule comprising an E. coli
CC strain 0157:H7 sequence selected from a clostridial cytotoxin-like gene,
CC a urease gene cluster, a RTX toxin-like gene cluster, a locus of
CC enterocyte effacement and 2 genes from its associated lymphocytic phage
CC 933W (a putative serine/threonine kinase and a tail fibre gene). E. coli
CC 0157:H7 can cause food poisoning, specifically acute haemorrhagic colitis
CC (which can develop into haemolytic uraemic syndrome). Also included are
CC an isolated DNA molecule comprising a nucleotide sequence identical to at
CC least 25 contiguous nucleotides contained in DNA sequences selected from
CC ACD18988-ACD19242 (being 255 E.coli 0157 DNA sequences which are not
CC found in E.coli K12), a recombinant DNA construction comprising the DNA
CC above and a method for detecting E. coli 0157:H7 (ATCC 43895) in a sample
CC (or distinguishing between 0157 and K12) using a probe derived from one
CC of the 255 sequences. The DNA sequences are useful in detecting E. coli
CC O157:H7 in a sample, for the early diagnosis of humans and livestock
CC infected with O157:H7, and in designing diagnostic probes which can be
CC used to distinguish strain O157:H7 from strain K12 using molecular
CC techniques. The present sequence is one of the 255 E. coli 0157:H7 DNA
CC sequence (termed OZID_1-OZID255). Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030023075 (Updated on 27-OCT-2003
CC to standardise OS field)
XX
SQ Sequence 14187 BP; 4451 A; 2859 C; 2282 G; 4595 T; 0 U; 0 Other;

```
  Query Match           100.0%;  Score 666;  DB 9;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-165;
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 786

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        785 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 726

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        725 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 666

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        665 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 606

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 546

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 486

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 426

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 366

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 306

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 246
```

```
Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 186

Qy        661 ATTTAA 666
              ||||||
Db        185 ATTTAA 180


RESULT 5
ADA29465
ID   ADA29465 standard; DNA; 687 BP.
XX
AC   ADA29465;
XX
DT   20-NOV-2003  (first entry)
XX
DE   DNA encoding Acinetobacter baumannii protein #752.
XX
KW   ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
KW   vaccine; plant biocontrol agent.
XX
OS   Acinetobacter baumannii.
XX
PN   US6562958-B1.
XX
PD   13-MAY-2003.
XX
PF   04-JUN-1999;   99US-00328352.
XX
PR   09-JUN-1998;   98US-0088701P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Breton G,  Bush D;
XX
DR   WPI; 2003-576092/54.
DR   P-PSDB; ADA33591.
XX
PT   New Acinetobacter baumanii proteins and nucleic acids, useful as reagents
PT   for diagnosing a bacterial disease, as components of antibacterial
PT   vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT   plants.
XX
PS   Example; SEQ ID NO 752; 328pp; English.
XX
CC   The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC   The A. baumannii nucleic acids and polypeptides are useful as reagents
CC   for diagnosing a bacterial disease, as components of antibacterial
CC   vaccines, as targets for antibacterial drugs, to detect the presence of
CC   A. baumannii and other Acinetobacter species in a sample, in screening
CC   compounds for the ability to interfere with the A. baumannii life cycle
CC   or to inhibit A. baumannii infection, and as biocontrol agents for
CC   plants. The present sequence represents DNA encoding an A. baumannii
CC   protein.
XX
SQ   Sequence 687 BP; 215 A; 102 C; 155 G; 215 T; 0 U; 0 Other;

  Query Match             32.8%;  Score 218.2;  DB 9;  Length 687;
  Best Local Similarity   60.1%;  Pred. No. 1.6e-47;
  Matches  383;  Conservative    0;  Mismatches  248;  Indels    6;  Gaps    1;

Qy         12 TGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAAT 71
              |||| ||| ||||||  |  ||| |||||  |||   || ||  |  |   ||| ||  |
Db         51 TGGTGTTTATGGTGCCAGTGGTTTTGGTAAGGAAGTTATGCCTTTAGTGCGTCAGCAGTT 110

Qy         72 AAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAA 131
                    |  | ||       | | |||      |||| | |||||||   |  | ||||
Db        111 TCCAACATTAAGTAAAGAACAATTTGC------TTTTATTGATGATGGACTATCTGGAAC 164

Qy        132 GAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAA 191
              ||   | ||||| |    ||| || ||      |  |||| | |     |||  |    |

Db        165 GACTTTAAATGGATACCCAGTACTATCCTATCTCGACTTTATCAGTAAACCTGCTGATCA 224

Qy        192 AAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAAT 251
              |||  | | |   | ||||| || |||  |  | | ||  |||   |   |       |
Db        225 TAAGGCTGTGACAATCGCTATAGCAAATAGTGTGGTTCGTGAGAAGTTAGTTAGCCTTTT 284

Qy        252 ATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCA 311
              |  |  | | || ||  |||   || || |  ||  | ||||||  ||  ||  |||| |
Db        285 AGAAAAAGATGGTGTACAACACCTAGCTGTTCAATCTACAAATACAGTAATTCTTGATGA 344

Qy        312 TACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCA 371
                 |  |||||||   ||      |  ||| ||||||  |    || || ||||||   |
Db        345 GGTTGAGATAGGTGAAGGAAGCTTACTTTGTCCCTTTACTTGCTTAACTTCTAATATAAA 404

Qy        372 TATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGG 431
              || || | |||||| || || || || ||    || || || |||||||||  ||||||
Db        405 AATTGGAAAGTTTTTCCACGCTAATATTTATAGTTATGTGGCTCATGATTGTGTAATAGG 464

Qy        432 AGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAA 491
              ||| |||||||||| || ||||| ||||| ||||| ||  || ||  |||||||||  |
Db        465 AGATTATGTTACATTCGCGCCTGGCGCTAAGTGTAACGGTAATATTCATATTGAAGATCA 524

Qy        492 TGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTAT 551
              ||| || || ||| | || ||||| ||||| |||||    ||| ||   ||  | || ||
Db        525 TGCCTACATTGGCACTGGAGCAGTTATTAAACAGGGCACACCTGATAAGCCTTTAATAAT 584

Qy        552 TGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTTCCTGCCGGTAT 611
              |||     || || ||| | || ||||| ||||| || || |||||||| || || ||  |
Db        585 TGGTAAAGGTGCTATTGTTGGAATGGGCGCTGTAGTAACGAAAAGTGTGCCAGCTGGCGT 644

Qy        612 AACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGA 648
              ||||||    || ||||| ||||||    |  |||||
Db        645 AACTGTTGTAGGTAATCCTGCAAGAATTTTAGAAAGA 681


RESULT 6
AEC31451
ID   AEC31451 standard; DNA; 14256 BP.
XX
AC   AEC31451;
XX
DT   20-OCT-2005  (first entry)
XX
DE   E. coli type O49 O-antigen gene cluster.
XX
KW   O-antigen; gene expression; DNA detection; gene; ds.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1138. .2223
FT                   /*tag=  a
FT                   /note= "Orf1"
FT   CDS             2223. .3122
FT                   /*tag=  b
FT                   /note= "Orf2"
FT   CDS             3180. .4058
FT                   /*tag=  c
FT                   /note= "Orf3"
FT   CDS             4063. .4602
FT                   /*tag=  d
FT                   /note= "Orf4"
FT   CDS             4613. .5728
FT                   /*tag=  e
FT                   /note= "Orf5"
FT   CDS             5733. .6386
FT                   /*tag=  f
FT                   /note= "Orf6. Specifically claimed in Claim 3"
FT   CDS             6388. .6804
FT                   /*tag=  g
FT                   /note= "Orf7"
FT   CDS             6848. .8350
FT                   /*tag=  h
```

```
FT                   /note= "Orf8"
FT   CDS             8347. .9579
FT                   /*tag=  i
FT                   /note= "Orf9. Specifically claimed in Claim 3"
FT   CDS             9551. .10723
FT                   /*tag=  j
FT                   /note= "Orf10"
FT   CDS             10738. .11559
FT                   /*tag=  k
FT                   /note= "Orf11. Specifically claimed in Claim 3"
FT   CDS             11556. .12455
FT                   /*tag=  l
FT                   /note= "Orf12. Specifically claimed in Claim 3"
FT   CDS             12586. .13557
FT                   /*tag=  m
FT                   /note= "Orf13"
XX
PN   CN1563044-A.
XX
PD   12-JAN-2005.
XX
PF   19-APR-2004; 2004CN-00019029.
XX
PR   19-APR-2004; 2004CN-00019029.
XX
PA   (TIAN-) TIANJIN BIOCHIP TECH CO LTD.
XX
PI   Wang L,  Yang J,  Feng L;
XX
DR   WPI; 2005-307119/32.
XX
PT   Nucleotide peculiar to 0-antigen of 049 type bacillus coli.
XX
PS   Claim 1; SEQ ID NO 1; 36pp; Chinese.
XX
CC   The invention relates to a nucleotide sequence (SEQ ID No:1) which is
CC   specific to the O-antigen of Escherichia coli O49. The nucleotide
CC   sequence represents the complete sequence (14256bp) of a gene cluster for
CC   controlling the synthesis of O-antigen in Escherichia coli O49. The
CC   nucleotide sequence can have one or several inserted, deleted or
CC   substituted bases, whilst still retaining its function. Also described is
CC   a PCR primer for the glycosyltransferase (galF) gene and a PCR primer for
CC   the oligosaccharide unit treatment (gnd) gene, both of which are
CC   contained within the E. coli type O49 O-antigen gene cluster. The
CC   invention also discloses a method for detecting and identifying E. coli
CC   type O49 in humans. This sequence represents the E. coli type O49 O-
CC   antigen gene cluster of the invention.
XX
SQ   Sequence 14256 BP; 4415 A; 2264 C; 2951 G; 4626 T; 0 U; 0 Other;

  Query Match           29.2%;  Score 194.4;  DB 14;  Length 14256;
  Best Local Similarity  58.3%;  Pred. No. 8.5e-41;
  Matches  367;  Conservative    0;  Mismatches  251;  Indels   12;  Gaps    1;

Qy         12 TGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAAT 71
              |||||||| |||||||||  |||  ||||||||      |||| ||| |||  |||||  |
Db       5747 TGGTATTTATGGTGCTGGTGGTTTCGGTAGAGAGGTGCTACCTATTATAAGACAACAGGT 5806

Qy         72 AAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAA 131
                  ||| |   | |  | | |||||||| | | |||||  | |   || |||
Db       5807 ------------TGGATTAAAATCAGATGTGGTTTTTATCGTTGATGATATCAAAGAAAA 5854

Qy        132 GAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAA 191
              |  ||| ||||      | ||  |  ||     |    |||               || |  |
Db       5855 GGTAGTCAATGAAACAAAGGTTGTAACATTTGATGAGTTTTGTTCGCTAAAAGATATGGA 5914

Qy        192 AAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAAT 251
              || |||| || |||||| || ||  |   |||  |  ||  |    |   |     |
Db       5915 TAAATATTATACTGTTGCAATAGCAGACAGTAAAGTGAGAATGGAGATAGATATAAAATG 5974

Qy        252 ATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCA 311
                |  | | ||  || | || ||  ||||  |   || ||| || ||  |    ||| |
Db       5975 CATCCGAGAGGGAATTAAGCCCATTTCTATTTATTCTCAAAACAGTGTGATAATGGATGA 6034

Qy        312 TACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCA 371
              |    |||| |||  ||| || ||| | ||||| |||  |   ||| |  | |||
Db       6035 TGTACTGATTGGTGATGGTGCGATTTTATCTCCTTTTACTTGTATAACATCAAATGTAAG 6094

Qy        372 TATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGG 431
               || ||||  | ||||||||| ||  ||||    ||  | || |||||||||   || ||
Db       6095 AATCGGGAAATCTTTTCATGCCAATTTATATAGTTATATAGCGCATGATTGTGTTATTGG 6154

Qy        432 AGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAA 491
              |||||||||  ||  ||  ||  ||  ||  |  ||||||||  ||  |||||  ||||  || ||  |
Db       6155 AGACTATGTCACTTTAGCACCAGGTGTAAAATGTAACGGTAATGTTATTATAGAGGATCA 6214

Qy        492 TGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTAT 551
              |||  |||  |  ||   |  ||  |||  |||||    |  ||         |     |||||  | ||
Db       6215 TGCTTATGTCGGAACTGGGGCAATAATTCGACCTGGGACCAAAACAAAACCACTCGTAAT 6274

Qy        552 TGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTTCCTGCCGGTAT 611
              |||          | |||||  ||  ||||||||  ||||||||||||||   ||  |||  |  ||  |
Db       6275 TGGAAGCCATTCTATTATTGGAATGGGGGCGGTTGTCACTAAAACAGTACCTTCAGGAAA 6334

Qy        612 AACTGTGTGCGGAAATCCAGCAAGAGAAAT 641
              |||  ||  |   ||  |||||  ||||  |   |  |
Db       6335 AACAGTATTTGGGAATCCGGCAAAAATATT 6364


RESULT 7
ACA47793
ID   ACA47793 standard; DNA; 1199 BP.
XX
AC   ACA47793;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Prokaryotic essential gene #29450.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Staphylococcus haemolyticus.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU43923.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 35663; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
```

```
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC   prokaryotic essential genes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1199 BP; 441 A; 178 C; 226 G; 354 T; 0 U; 0 Other;

  Query Match             10.7%;  Score 71.4;  DB 8;  Length 1199;
  Best Local Similarity   48.6%;  Pred. No. 1.4e-08;
  Matches  195;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy        194 AGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAATAT 253
              | |  |||| | |||||||||  | | || | || || || | | | | |   | ||
Db        769 ATTCATTTATTATTGCTATTGGAAGTAATCATATTCGTCAAAAACTTTTTCATAAATTAG 828

Qy        254 TATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATA 313
              | |  |    | | ||     |||  |||  |||||||     |||||   |  |  |
Db        829 AAGTTAAAATTGATCAATATGCAACATTAATTCATCCAACAGCAGTTGTGAGTCCTAGTG 888

Qy        314 CTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATA 373
              | | ||| |||   ||  |  || |    ||   ||     |||  | |  |  ||   |
Db        889 CGAAGATTGGTTACGGAACGGTTGTAATGCCTAATGCGGTGATAAATGCGGACTCTATAA 948

Qy        374 TAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAG 433
              | ||| | |  | |       |||| |  | ||   || | ||| |||  |||||| || |
Db        949 TTGGAAAGCATGTCATAATCAACACAGGCGCAATAGTCGAACACGATAATCAAATTGGTG 1008

Qy        434 ACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATG 493
              | ||||||    || | |||   |||| ||     ||   ||||    | |  ||
Db       1009 ATTATGTTCATATTTCACCTAATGCTACATTAGCAGGTGGTGTTAAAGTAGGCGAGGCAT 1068

Qy        494 CATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTATTG 553
              || | || |       ||| |   | |  | |    ||   | | |   |     |||||
Db       1069 CACACATTGCAATTAATGCTGCTGTAATACCGTTGATTGAAATTGGTGAAAGATGTATTG 1128

Qy        554 GCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAA 594
                |  | ||     | ||  ||      ||||   ||  ||
Db       1129 TGGGAGCAGGTGCAACAGTCATAAATAATGTTAAAACAGAA 1169


RESULT 8
AEB42401_32/c
Continuation (33 of 34) of AEB42401 from base 3200001 (L. pneumophila DNA SEQ ID NO 6733
WP Sequence split into 34 fragments  LOCUS AEB42401 Accession Aeb42401
WP     Fragment Name      Begin     End
WP     AEB42401_00            1    110000
WP     AEB42401_01       100001    210000
WP     AEB42401_02       200001    310000
WP     AEB42401_03       300001    410000
WP     AEB42401_04       400001    510000
WP     AEB42401_05       500001    610000
WP     AEB42401_06       600001    710000
WP     AEB42401_07       700001    810000
WP     AEB42401_08       800001    910000
WP     AEB42401_09       900001   1010000
WP     AEB42401_10      1000001   1110000
WP     AEB42401_11      1100001   1210000
WP     AEB42401_12      1200001   1310000
WP     AEB42401_13      1300001   1410000
WP     AEB42401_14      1400001   1510000
WP     AEB42401_15      1500001   1610000
WP     AEB42401_16      1600001   1710000
WP     AEB42401_17      1700001   1810000
WP     AEB42401_18      1800001   1910000
WP     AEB42401_19      1900001   2010000
WP     AEB42401_20      2000001   2110000
WP     AEB42401_21      2100001   2210000
WP     AEB42401_22      2200001   2310000
WP     AEB42401_23      2300001   2410000
WP     AEB42401_24      2400001   2510000
WP     AEB42401_25      2500001   2610000
WP     AEB42401_26      2600001   2710000
WP     AEB42401_27      2700001   2810000
WP     AEB42401_28      2800001   2910000
WP     AEB42401_29      2900001   3010000
WP     AEB42401_30      3000001   3110000
WP     AEB42401_31      3100001   3210000
WP     AEB42401_32      3200001   3310000
WP     AEB42401_33      3300001   3345687

  Query Match              7.1%;  Score 47.6;  DB 14;  Length 110000;
  Best Local Similarity   46.4%;  Pred. No. 0.11;
  Matches  155;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy        282 AAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTC 341
              ||||  | |   |    ||| | |  || ||| |||   ||    ||   ||||||
Db      80394 AAAAAGTACGGTTTTTATTGGTAACAATAATATTATTCGAGAATATGTGACTATTAACCG 80335

Qy        342 TCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATA 401
              || |    |       | || ||| | || |    |  |      |  | |  |
Db      80334 AGGAACTGCTGGCGGAGGAGGAGAAACGCATGTCGGTAATCATAATTTAATCATGACTTC 80275

Qy        402 CTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAA 461
                | || || || ||| |||||| ||| || ||    |   | |||  || || |
Db      80274 AGTCCATGTAGCCCACGATGTTCAAATGGGAAACAATAACATTATCGCTAATGCGGTTGC 80215

Qy        462 ATGTAATGGATATGTTGTTATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAA 521
              |   ||| |||| |||||||| ||    |   ||| ||| |       |  ||| |
Db      80214 CATTGGCGGACATGTAGTTATTGATGATTGGGTGACTATTGGCGCTTTGTGCGGGATTCA 80155

Qy        522 GCAGGGTGTTCCTAATCGCCCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGC 581
              ||   |||||      ||  | | | |||    | | |   |  | |  ||     |
Db      80154 TCAATTTGTTCAGTTAGGCAGAATGAGTATGATTGGTGCACAAAGCAAAATTACAAAAGA 80095

Qy        582 TGTTGTCACTAAAAGTGTTCCTGCCGGTATAACT 615
              |||| |  || | | | |   | | || |   ||
Db      80094 TGTTTTGCCTTATACTTTGGTTTCTGGAAATCCT 80061


RESULT 9
ADH88991
ID   ADH88991 standard; DNA; 624 BP.
XX
AC   ADH88991;
XX
DT   22-APR-2004  (first entry)
XX
DE   Escherichia coli  K1 S26 mutant coding sequence, SEQ ID 25.
XX
KW   Pathogenic microorganism; K1 S26; mutant; pathogenicity; virulence;
KW   bacterial infection; extra-intestinal infection; enterobacterium; gene;
KW   ds.
XX
OS   Escherichia coli.
OS   Synthetic.
```

```
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .624
FT                   /*tag=  a
FT                   /product= "K1 S26 mutant"
XX
PN   WO2004005535-A2.
XX
PD   15-JAN-2004.
XX
PF   09-JUL-2003; 2003WO-EP008209.
XX
PR   09-JUL-2002; 2002FR-00008636.
XX
PA   (MUTA-) MUTABILIS SA.
XX
PI   Escaich S;
XX
DR   WPI; 2004-091381/09.
DR   P-PSDB; ADH88976.
XX
PT   Identifying and selecting a gene required for the proliferation in vivo
PT   of a pathogenic microorganism comprises determining the virulence of
PT   mutant genes on an experimental model of infection, and their effect on
PT   enteric colonization.
XX
PS   Claim 5; SEQ ID NO 25; 60pp; English.
XX
CC   The present invention relates to a method for identifying and selecting a
CC   gene required for the proliferation in vivo of a pathogenic
CC   microorganism. The method comprises determining the virulence of mutant
CC   genes on an experimental model of infection, and their effect on enteric
CC   colonization in an axenic mouse model. ADH88967-ADH88981 and ADH88982-
CC   ADH88996 are pathogenicity or virulence targets and their coding
CC   sequences which were used in the method of the invention. Compounds'
CC   capable of inhibiting pathogenicity or virulence target expression are
CC   useful for developing medicinal products for inhibiting a bacterial
CC   infection, in particular an extra-intestinal infection in the case of
CC   enterobacteria.
XX
SQ   Sequence 624 BP; 221 A; 72 C; 124 G; 207 T; 0 U; 0 Other;

  Query Match             6.9%;  Score 46.2;  DB 12;  Length 624;
  Best Local Similarity   48.0%;  Pred. No. 0.052;
  Matches  178;  Conservative    0;  Mismatches  178;  Indels   15;  Gaps    1;

Qy        273 AATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGC 332
              ||| |  || | | ||  || |    || | |      ||| |||  | |||  |||
Db        255 AATTAACATTATAGATAAAACTGCTATTCTATCACCAAATATTATACTGGGTGATGGAAT 314

Qy        333 TATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGC 392
              | |||||  |    |  || | |      | |||||   |||        ||
Db        315 TTTTATTGGTAAAATGTGTATACTTAACCGTGATACTAGAATACATGATGCCGTTGTAAT 374

Qy        393 AAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACATTTGCTCC 452
              |||  |       |  ||| |||||  |  | |||||||   ||  |   |   |  || |
Db        375 AAATACTAGGAGTTTAATTGAACATGGTAATGAAATAGGCTGCTGTAGCAATATCTCTAC 434

Qy        453 TGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATAGGCTCGGGTGC 512
              |   | |  |  ||||||  |||||  | ||| |||  |  | | | | ||
Db        435 TAATGTTGTACTTAATGGTGATGTTTCTGTTGGAGAAGAAACTTTTGTTGG--------- 485

Qy        513 AGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTATTGGCGCGGGAGCCATTATAGG 572
              ||    |  ||||   ||| |||   | |   | ||| |       |||| ||
Db        486 ------TAGCGTGACTGTTGTAAATGGCCAGTTGAAGCTAGGCTCAAAGAGTATTATTGG 539

Qy        573 TATGGGGGCTGTTGTCACTAAAAGTGTTCCTGCCGGTATAACTGTGTGCGGAAATCCAGC 632
              |   ||| | ||||| | || || | | ||      | |    ||    || | |||| |
Db        540 TTCTGGGTCGGTTGTAATTAGAAATATACCAAGTAATGTTGTAGTTGCTGGGACTCCAAC 599

Qy        633 AAGAGAAATGA 643
              ||||  ||| |
Db        600 AAGATTAATTA 610
```

```
RESULT 10
AAH65371
ID   AAH65371 standard; DNA; 645 BP.
XX
AC   AAH65371;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 406.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-00127688.
XX
PR   16-DEC-1999;   99JP-00377484.
PR   07-APR-2000; 2000JP-00159162.
PR   03-AUG-2000; 2000JP-00280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
DR   P-PSDB; AAG90152.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analyzing
PT   expression profile or pattern of a gene and identifying homologous gene.
XX
PS   Claim 8; SEQ ID NO 406; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and analysing
CC   the expression profile or expression pattern of a gene derived from
CC   Coryneform bacterium, and identifying a homologue of a gene derived from
CC   coryneform bacterium. Coryneform bacteria are useful for producing amino
CC   acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from the European Patent Office
XX
SQ   Sequence 645 BP; 164 A; 118 C; 180 G; 183 T; 0 U; 0 Other;

  Query Match             6.8%;  Score 45.2;  DB 5;  Length 645;
  Best Local Similarity   46.1%;  Pred. No. 0.096;
  Matches  152;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy        206 TTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGG 265
              |||| ||||  ||    | ||| || || |   | ||||       ||          ||
Db        230 TTGCGATTGGAAACCCAACGATTCGCCATAAGCTTTCTGCTCTTTTAGACAAGTCTGGTT 289

Qy        266 TTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTA 325
              | ||  || | ||        |||||   |   ||  |         ||     || | ||
Db        290 TCGATGCATTTACGGCGGTGCATCCCTCTGCAGTGATAGGAAGCCAAGTAGTGTTTGGGT 349

Qy        326 GTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTT 385
              |||  ||| || | | | |     |      || || || |||        | ||   |  |
Db        350 CTGGGGCTGTTGTATGTTCTGGGGCCGTGATTTCGACCAATGTAAGGCTTGGTGTGCATG 409

Qy        386 TTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACAT 445
```

```
Db        410 TGCATATCAACCCAAATGTGACAATCGGGCATGATTCAATTCTTGAAGACTATGTTTCGA 469

Qy        446 TTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATAGGCT 505

Db        470 TAAATCCGGCTGCAACAATTTCGGGTGAAGTCCTGATCATGGAAGAAACTCTTATTGGAG 529

Qy        506 CGGGTGCAGTAATTAAGCAGGGTGTTCCTA 535

Db        530 CTGCGGCTACAGTATTGCAAGGACTGACTA 559
```

RESULT 11
ACA02087
ID ACA02087 standard; DNA; 645 BP.
XX
AC ACA02087;
XX
DT 04-JUN-2003 (first entry)
XX
DE C. glutamicum derived ORF SEQ ID 2078.
XX
KW Coryneform; nucleic acid array; fermentation; culture; ds.
XX
OS Corynebacterium glutamicum.
XX
PN DE10128510-A1.
XX
PD 19-DEC-2002.
XX
PF 13-JUN-2001; 2001DE-01028510.
XX
PR 13-JUN-2001; 2001DE-01028510.
XX
PA (DEGS) DEGUSSA AG.
XX
PI Farwick M, Moeckel B, Pfefferle W, Bathe B, Huthmacher K;
XX
DR WPI; 2003-279970/28.
XX
PT New nucleic acid array useful for monitoring mRNA expression of
PT Corynebacterium glutamicum during fermentation, comprising nucleic acid
PT from Corynebacterium glutamicum.
XX
PS Claim 1; Page 682; 709pp; German.
XX
CC This invention describes a novel nucleic acid array involving
CC Corynebacterium glutamicum polynucleotides. The arrays are used to
CC analyse C. glutamicum, particularly for monitoring a fermentation process
CC to determine expression levels of C. glutamicum cellular mRNA. Such
CC monitoring particularly differentiates between expression levels of
CC different strains of C. glutamicum and allows the adjustment of different
CC culture and fermentation conditions. ACA00010-ACA02188 represent C.
CC glutamicum derived polynucleotides described in the disclosure of the
CC invention
XX
SQ Sequence 645 BP; 164 A; 118 C; 180 G; 183 T; 0 U; 0 Other;

Query Match 6.8%; Score 45.2; DB 8; Length 645;
Best Local Similarity 46.1%; Pred. No. 0.096;
Matches 152; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

```
Qy        206 TTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGG 265

Db        230 TTGCGATTGGAAACCCAACGATTCGCCATAAGCTTTCTGCTCTTTTAGACAAGTCTGGTT 289

Qy        266 TTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTA 325

Db        290 TCGATGCATTTACGGCGGTGCATCCCTCTGCAGTGATAGGAAGCCAAGTAGTGTTTGGGT 349

Qy        326 GTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTT 385

Db        350 CTGGGGCTGTTGTATGTTCTGGGGCCGTGATTTCGACCAATGTAAGGCTTGGTGTGCATG 409
```

```
Qy        386 TTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACAT 445

Db        410 TGCATATCAACCCAAATGTGACAATCGGGCATGATTCAATTCTTGAAGACTATGTTTCGA 469

Qy        446 TTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATAGGCT 505

Db        470 TAAATCCGGCTGCAACAATTTCGGGTGAAGTCCTGATCATGGAAGAAACTCTTATTGGAG 529

Qy        506 CGGGTGCAGTAATTAAGCAGGGTGTTCCTA 535

Db        530 CTGCGGCTACAGTATTGCAAGGACTGACTA 559
```

RESULT 12
AAH68525
ID AAH68525 standard; DNA; 349980 BP.
XX
AC AAH68525;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 7060.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-00127688.
XX
PR 16-DEC-1999; 99JP-00377484.
PR 07-APR-2000; 2000JP-00159162.
PR 03-AUG-2000; 2000JP-00280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
PT expression profile or pattern of a gene and identifying homologous gene.
XX
PS Disclosure; SEQ ID NO 7060; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and analysing
CC the expression profile or expression pattern of a gene derived from
CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the European Patent Office
XX
SQ Sequence 349980 BP; 80289 A; 91081 C; 97378 G; 81232 T; 0 U; 0 Other;

Query Match 6.8%; Score 45.2; DB 5; Length 349980;
Best Local Similarity 46.1%; Pred. No. 0.7;
Matches 152; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

```
Qy        206 TTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGG 265
```

```
Db      86785 TTGCGATTGGAAACCCAACGATTCGCCATAAGCTTTCTGCTCTTTTAGACAAGTCTGGTT 86844

Qy        266 TTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTA 325

Db      86845 TCGATGCATTTACGGCGGTGCATCCCTCTGCAGTGATAGGAAGCCAAGTAGTGTTTGGGT 86904

Qy        326 GTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTT 385

Db      86905 CTGGGGCTGTTGTATGTTCTGGGGCCGTGATTTCGACCAATGTAAGGCTTGGTGTGCATG 86964

Qy        386 TTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACAT 445

Db      86965 TGCATATCAACCCAAATGTGACAATCGGGCATGATTCAATTCTTGAAGACTATGTTTCGA 87024

Qy        446 TTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATAGGCT 505

Db      87025 TAAATCCGGCTGCAACAATTTCGGGTGAAGTCCTGATCATGGAAGAAACTCTTATTGGAG 87084

Qy        506 CGGGTGCAGTAATTAAGCAGGGTGTTCCTA 535

Db      87085 CTGCGGCTACAGTATTGCAAGGACTGACTA 87114
```

RESULT 13
ACF74035
ID ACF74035 standard; DNA; 609 BP.
XX
AC ACF74035;
XX
DT 20-NOV-2003 (first entry)
XX
DE Staphylococcus aureus DNA #1715.
XX
KW Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;
KW enzymatic assay; antibiotic target; gene; ds.
XX
OS Staphylococcus aureus.
XX
PN WO200294868-A2.
XX
PD 28-NOV-2002.
XX
PF 27-MAR-2002; 2002WO-IB002637.
XX
PR 27-MAR-2001; 2001GB-00007661.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Masignani V, Mora M, Scarselli M;
XX
DR WPI; 2003-120786/11.
DR P-PSDB; ABM72475.
XX
PT New Staphylococcus aureus protein, useful as a vaccine for treating or
PT preventing Staphylococcal infection, specifically an infection caused by
PT S. aureus, e.g. sepsis.
XX
PS Claim 6; SEQ ID NO 3429; 49pp; English.
XX
CC The invention relates to novel genes and encoded proteins from
CC Staphylococcus aureus. A composition comprising the S. aureus protein, a
CC nucleic acid encoding the protein, or an antibody to the protein, is
CC useful as a pharmaceutical, particularly as a vaccine for treating or
CC preventing infection due to Staphylococcus bacteria, specifically an
CC infection caused by S. aureus. The composition is particularly useful for
CC treating or preventing sepsis in a patient. The composition can also be
CC used for diagnostics. The protein is also used in an assay for enzymatic
CC studies and as a target for antibiotics. This sequence represents one of
CC the novel S. aureus genes of the invention
XX
SQ Sequence 609 BP; 217 A; 89 C; 132 G; 171 T; 0 U; 0 Other;

Query Match 6.7%; Score 44.6; DB 8; Length 609;

Best Local Similarity 54.6%; Pred. No. 0.14;
Matches 89; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```
Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484

Db        326 ATATAGGAGACAATGTTTTAATAGCAGCCGGTGCGAAAGTTTTAGGAAATATTAAAATAA 385

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544

Db        386 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 445

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587

Db        446 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 488
```

RESULT 14
AAS51946
ID AAS51946 standard; DNA; 627 BP.
XX
AC AAS51946;
XX
DT 13-FEB-2002 (first entry)
XX
DE Staphylococcus aureus DNA for cellular proliferation protein #363.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
KW antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU34087.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Claim 27; SEQ ID NO 4528; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence encodes an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent

```
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 627 BP; 228 A; 88 C; 138 G; 173 T; 0 U; 0 Other;

  Query Match             6.7%;  Score 44.6;  DB 4;  Length 627;
  Best Local Similarity   54.6%;  Pred. No. 0.14;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              | |||||||||| |||||  | | ||  | || || ||| |   ||| || ||   ||
Db        356 ATATAGGAGACAATGTTTTAATAGCAGCGGGTGCGAAAGTTTTAGGAAATATTAAAATAA 415

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |    |||| | |||| ||  |   | ||||  ||   ||    ||||| | |    ||
Db        416 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 475

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587
                || ||||      ||    ||| | || |   | || |||
Db        476 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 518


RESULT 15
ACA20258
ID   ACA20258 standard; DNA; 642 BP.
XX
AC   ACA20258;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Prokaryotic essential gene #1915.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Staphylococcus aureus.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU16388.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 8128; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC   prokaryotic essential genes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 642 BP; 235 A; 90 C; 138 G; 179 T; 0 U; 0 Other;

  Query Match             6.7%;  Score 44.6;  DB 8;  Length 642;
  Best Local Similarity   54.6%;  Pred. No. 0.14;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              | |||||||||| |||||  | | ||  | || || ||| |   ||| || ||   ||
Db        356 ATATAGGAGACAATGTTTTAATAGCAGCCGGTGCGAAAGTTTTAGGAAATATTAAAATAA 415

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |    |||| | |||| ||  |   | ||||  ||   ||    ||||| | |    ||
Db        416 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 475

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587
                || ||||      ||    ||| | || |   | || |||
Db        476 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 518


Search completed: May 10, 2006, 10:59:23
Job time : 314.898 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 11:00:02 ; Search time 2133.07 Seconds
(without alignments)
14608.170 Million cell updates/sec

Title: US-09-423-093B-2_COPY_13156_13821
Perfect score: 666
Sequence: 1 atgaatttgtatggtatttt.....gatcgccaacatctatttaa 666

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	57.6	8.6	1101	10	CNS0039G	AL063921 Drosophil
	2	48	7.2	905	8	CX112330	CX112330 EI075F18
	3	46.8	7.0	466	1	AW333225	AW333225 S19A1 AGS
	4	46.4	7.0	928	10	CNS00DKY	AL071865 Drosophil
c	5	46.2	6.9	1101	10	CNS00EVL	AL069706 Drosophil
	6	46	6.9	617	8	CX098326	CX098326 EHAHK13TR
	7	46	6.9	622	8	CX099301	CX099301 EHAHY22TR
	8	46	6.9	633	8	CX086360	CX086360 EHACI92TR
	9	46	6.9	633	8	CX093673	CX093673 EHAFO42TR
	10	46	6.9	635	8	CX083169	CX083169 EHAB786TR
c	11	46	6.9	642	8	CX083304	CX083304 EHAB975TR
	12	46	6.9	663	8	CX091008	CX091008 EHAEM45TR
	13	45.8	6.9	600	1	AU037922	AU037922 AU037922
	14	45.8	6.9	1101	10	CNS000D1	AL065414 Drosophil
	15	45.8	6.9	1295	10	CL467398	CL467398 SAIL_1269
c	16	45.6	6.8	1200	1	AJ928743	AJ928743 AJ928743
c	17	45.4	6.8	576	10	CZ533139	CZ533139 SRAA-aac8
	18	45.4	6.8	987	10	CNS014PQ	AL104456 Drosophil
c	19	45.2	6.8	566	1	AU270659	AU270659 AU270659
c	20	44.6	6.7	748	10	BX246108	BX246108 Danio rer
c	21	44.6	6.7	936	10	DU002907	DU002907 285322 To
c	22	44.6	6.7	1092	10	CNS020K7	AL175696 Tetraodon
c	23	44.4	6.7	596	9	AZ819906	AZ819906 2M0091L16
	24	44.4	6.7	639	8	CX083039	CX083039 EHAB585TR
	25	44.4	6.7	642	8	CX093521	CX093521 EHAFM23TR
c	26	44	6.6	565	9	BH707890	BH707890 BOMHH28TR
c	27	44	6.6	987	10	CNS014PQ	AL104456 Drosophil
	28	43.8	6.6	519	1	AW173150	AW173150 xj84a03.x
c	29	43.8	6.6	670	8	CX291390	CX291390 C03007D09
	30	43.6	6.5	579	2	BF199989	BF199989 WHE2251_F
	31	43.4	6.5	552	6	CF664918	CF664918 RTCNT1_12
c	32	43.2	6.5	497	1	AU270658	AU270658 AU270658
	33	43.2	6.5	647	7	CN442953	CN442953 re56b10.y
	34	43.2	6.5	823	10	CW683954	CW683954 OG_BBa004
c	35	43	6.5	703	10	AG376401	AG376401 Mus muscu
c	36	43	6.5	840	8	DN051340	DN051340 JGI_CABA6
	37	43	6.5	948	11	CNS04ORW	AL300389 Tetraodon
c	38	43	6.5	1029	10	CNS00ZNT	AL097907 Drosophil
	39	43	6.5	1086	10	CNS00YXK	AL096962 Drosophil
	40	43	6.5	1101	10	CNS00LT2	AL078714 Drosophil
	41	42.8	6.4	405	1	AU261540	AU261540 AU261540
	42	42.8	6.4	820	6	CA767431	CA767431 AF53-Rpf_
	43	42.6	6.4	397	9	BZ767975	BZ767975 SALK_1396
	44	42.4	6.4	531	1	AW611539	AW611539 hg85b10.x
	45	42.4	6.4	912	10	CG939336	CG939336 MBEJO95TR

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

```
  Query Match             8.6%;  Score 57.6;  DB 10;  Length 1101;
  Best Local Similarity   17.1%;  Pred. No. 0.00069;
  Matches   64;  Conservative  173;  Mismatches  137;  Indels    0;  Gaps    0;

Qy         14 GTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAATAA 73
              :::|: |::::: :: :: ::|:| : :| :|: | ::     :   :::| :| |::::|
Db        902 DDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTWDAAADDWWA 843

Qy         74 AGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAAGA 133
              :: :: :|::::: : :|| ||:: :  ::  :  | ::: :::: ::::: |::::: |
Db        842 DDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRDAADDRDDAA 783

Qy        134 AAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAA 193
               : |:  |   :|::: : |:| :|  :  :: ::: ::|     : :|::::: :
Db        782 TWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTWKRRWKR 723

Qy        194 AGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAATAT 253
              ::|:::: :|::| : ::  : ::  :|   : |  :: ::::| ::  : :   ||::|
Db        722 RDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRDRATWDRTDAWWADAAWWT 663

Qy        254 TATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATA 313
              | :|: :  :::: ::   :: :: || ||:: :|  |:::|: :::   :::::: |:
Db        662 TTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRADRWDRWAADT 603

Qy        314 CTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATA 373
               |: :  ::::|: :  :  :: |          :||: :   ||:| :|  ||:: : |:|
Db        602 WTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTTATWTTWAARAAWWAWWAWA 543

Qy        374 TAGGGAGGTTTTTT 387
              |    :  |:|||:
Db        542 TTTATWTTTWTTTW 529


RESULT 2
CX112330
LOCUS       CX112330                905 bp    mRNA    linear   EST 03-JUN-2005
DEFINITION  EI075F18 A normalized whole-life-cycle cDNA library of rice Oryza
            sativa (indica cultivar-group) cDNA clone EI075F18 5', mRNA
            sequence.
ACCESSION   CX112330
VERSION     CX112330.1  GI:66925482
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 905)
  AUTHORS   Zhang,J., Feng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D.,
            Han,B., Zhang,Q. and Wang,S.
  TITLE     Features of the expressed sequences revealed by a large-scale
            analysis of ESTs from a normalized cDNA library of the elite indica
            rice cultivar Minghui 63
  JOURNAL   Plant J. 42 (5), 772-780 (2005)
   PUBMED   15918889
COMMENT     Contact: Wang S
            National Key Laboratory of Crop Genetic Improvement
            Hauzhong Agricultural University
            Wuhan 430070, China
            Tel: 86-27-87282044
            Fax: 86-27-87287092
            Email: shipingwang@hotmail.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .905
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /strain="indica"
                     /cultivar="Minghui 63"
                     /db_xref="taxon:39946"
                     /clone="EI075F18"
                     /tissue_type="whole plant"
                     /dev_stage="whole-life-cycle"
                     /lab_host="E. coli DH10B"
                     /clone_lib="A normalized whole-life-cycle cDNA library of
                     rice"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
                     library is constructed based on the strategy of satuation
                     hybridization with genomic DNA using rice cultivar Minghui
                     63. This library consists of cDNA from 15 directionally
                     cloned cDNA libraries constructed with different tissues
                     from 9 developmental stages."
ORIGIN

  Query Match             7.2%;  Score 48;  DB 8;  Length 905;
  Best Local Similarity   46.5%;  Pred. No. 0.21;
  Matches  153;  Conservative    0;  Mismatches  176;  Indels    0;  Gaps    0;

Qy        284 AACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTC 343
              || |||  ||         ||  | |  ||| |  |||  |       |  |  |||| ||
Db        376 AAGATCTCAAGTATAAGATTGGTAACAATAATGAGATCAGAGAGTACTGTTCTATTCATC 435

Qy        344 CCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACT 403
              || |  || |   | |     ||  ||| ||     |  || |  ||    |  |
Db        436 GCTCTTCTAAATCTTGTGACTGCACGGTTATTGGTGACAATAATCTTATAATGGGTTCAT 495

Qy        404 CATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAAT 463
                |  |||||||||||     || ||  || ||    | |||||||       ||  ||
Db        496 GCCATATAGCACATGATTGCAGGATTGGCAACAATAACATATTTGCTAACAATACTCTAT 555

Qy        464 GTAATGGATATGTTGTTATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGC 523
              |  |||  |||||||  |||||||     |  |||  |     || ||  ||  |  | |
Db        556 TTGCTGGCCATGTTGTAGTTGAAGATTGCACCCATACTGCAGGGGCTGTTGTTGTCCATC 615

Qy        524 AGGGTGTTCCTAATCGCCCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTG 583
              |   |  || || | |  || ||    ||   |  |||| |  | |          | ||
Db        616 AATTTTGTCATATTGGGTCATTTTCATTTCTAGGTGGAGGCTCTGTGATCGCACAAGATG 675

Qy        584 TTGTCACTAAAAGTGTTCCTGCCGGTATA 612
              | |  |     |  | |  ||| |  |||
Db        676 TGGCAAGATTCATGGATGGTGCNGTGATA 704


RESULT 3
AW333225
LOCUS       AW333225                466 bp    mRNA    linear   EST 31-JAN-2000
DEFINITION  S19A1 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
ACCESSION   AW333225
VERSION     AW333225.1  GI:6829582
KEYWORDS    EST.
SOURCE      Pneumocystis carinii
  ORGANISM  Pneumocystis carinii
            Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
            Pneumocystidaceae; Pneumocystis.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Smulian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,
            Edman,J.C., Kovacs,J. and Cushion,M.
  TITLE     Expressed sequence tags from Pneumocystis carinii
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Staben C
            School of Biological Sciences
            University of Kentucky
            101 Morgan Building, University of Kentucky, Lexington, KY
            40506-0225, USA
            Tel: 606 257 2161
            Fax: 606 257 1717
            Email: staben@pop.uky.edu.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Pneumocystis carinii"
                     /mol_type="mRNA"
                     /db_xref="taxon:4754"
                     /lab_host="E. coli"
                     /clone_lib="AGS-1"
                     /note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI;
                     P. carinii organisms (3x10e9) from a single rat (99-1-6,
```

sacrificed on 3/17/99) at Cincinnati VA facilities. Trizol extracted RNA. Oligo dT priming, standard conditions described by vendor, Stratagene. Further details see www.uky.edu/Project/Pneumocystis/"

ORIGIN

Query Match 7.0%; Score 46.8; DB 1; Length 466;
Best Local Similarity 50.9%; Pred. No. 0.37;
Matches 111; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```
Qy        183 TTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTC 242
              ||   | | | | |||||| ||| |     |||| |||  || |             | |
Db          8 TTCAATTATACAATATTTTTATGATTAAGATGCTTATGGAAATAATATTATATCTTTTTT 67

Qy        243 TGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGT 302
              | ||  ||  |  ||  |  |   |||   || |  || |||  |||||||||     |
Db         68 TAAGAAAAATTGCTTTAAGAGTTATGATGAAAAATATAAAAATAATCCAAATACTTCTCA 127

Qy        303 TTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTAC 362
               ||   | | | | |||| || |  |  |||| ||      ||| |  ||| ||||||
Db        128 ATAATTTAAGATTTTTATAGATATTTTCTATATTTTTATAGAATTTATATCAAAATCTAC 187

Qy        363 TAATACTCATATAGGGAGGTTTTTTCATGCAAACATAT 400
              ||||||    |     |   ||||  ||  |   || |
Db        188 TAATACGTGAAATATTATAATTTTATATAAATTTATTT 225
```

RESULT 4
CNS00DKY
LOCUS CNS00DKY 928 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL071865
VERSION AL071865.1 GI:4948170
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 928)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .928
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR27A24"
/clone_lib="RPCI-98"
/note="end : T7"

ORIGIN

Query Match 7.0%; Score 46.4; DB 10; Length 928;
Best Local Similarity 26.9%; Pred. No. 0.54;
Matches 99; Conservative 92; Mismatches 177; Indels 0; Gaps 0;

```
Qy        187 TTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAG 246
              | |:| || ||| :||:|     | ||  :|:: | || : |          : | | : :
Db        489 TAAWATAAATATGWTAWTAAATTTTTTATKAWWAAAAAAWAATWTTTTTWWTTTTTWAWA 548

Qy        247 TCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTAT 306
                :  || :|        || :|    ||| |:|| :  |:  |||::   :  |||||
Db        549 AAWTATTTTWTTTTTTTTWTTTWAWTTTAATTWTATWTTAWTTAAAWWTAWWATTTTAT 608

Qy        307 GATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAAT 366
               :| |   ::  |||  |: |       :|:||: :   |:|  || |: |   |  :|
Db        609 AWTWAATAWWAAATATWTWATATAATATWTWTTWTWTAWTWTTATATAWAAAWAAAAWAA 668

Qy        367 ACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAA 426
              |   : |||   |  ||| :  :|  ||: | |   | :|   ||     |  :| | :
Db        669 AAAAWAATAAAAAATTTTAWAAWTAAAAWAAAAATTTAWTTTTTTTTTTTTTTWTATWTW 728

Qy        427 ATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAA 486
              |:|  : | ||| ::: : |:  :         ::::: ::::  :::   : ::::: ::
Db        729 AWATAWAAATATWWWWWWATWTWDGKNWNNNAWWWWWWWWWWAWWWWAAAWWWWWWWWW 788

Qy        487 GACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTT 546
              | ||   | | | :        :: :::::::::: |::::::  :  : : :   :|
Db        789 WAAAAAAAAAAAAAWAAAAAAAWDDDDDDWDDDWWKAAKKKKKKKKKKKKKKKBGKKKT 848

Qy        547 ATTATTGG 554
                |: ||:|
Db        849 KTKTTTKG 856
```

RESULT 5
CNS00EVL/c
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"

```
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             6.9%;  Score 46.2;  DB 10;  Length 1101;
  Best Local Similarity   30.4%;  Pred. No. 0.62;
  Matches  137;  Conservative   78;  Mismatches  235;  Indels    0;  Gaps    0;

Qy        138 TAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAAGTA 197
              :| |  ||:|  | |  | |  :  ::    |:| |  :|     :|||||| : :  |
Db        857 WATTTTTTWTWTATTWWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTWTWTT 798

Qy        198 TTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAATATTATT 257
              :| | :| |:  ||::  ::|| :: | | :        : | |   : | | |:||:||
Db        797 WTATTWTATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTTWTT 738

Qy        258 ACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTAT 317
              |  || |:  | : | :|| :  ||     |    || |||:|  | : | |
Db        737 TTATTWTTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTTWTATTTWAAATTA 678

Qy        318 GATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGG 377
              ||  || :      :||||||     :|  ||| | |||  |  | |    : | |
Db        677 TTTAATTAWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAAAAA 618

Qy        378 GAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTA 437
              | |||||  ||  | : :  |  | :      |  | :| :|     : ||  : | |:
Db        617 TATTTTTTTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTATTW 558

Qy        438 TGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATA 497
              : ||| |||   |  |    : |: :: :  |:|  :  ::::  ||   :::   |
Db        557 WTTTAAATTATWTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWTWTTT 498

Qy        498 TATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTATTGGCGC 557
              | :|       |  : | :: :|  : ::::::   :   |      |: :: ::||:|:
Db        497 TTWATTWATAAWTTTWTTTWWAWATTWWKKKKKKAWADTGAGARATKTDKKKKKKGGSGS 438

Qy        558 GGGAGCCATTATAGGTATGGGGGCTGTTGT 587
              ||| |          :   || ||  | |||
Db        437 GGGGGGGSNCNNNNSNGGGGNGGGGGGTGT 408


RESULT 6
CX098326
LOCUS       CX098326                617 bp    mRNA    linear   EST 15-DEC-2004
DEFINITION  EHAHK13TR E. histolytica Normalized cDNA library Entamoeba
            histolytica cDNA, mRNA sequence.
ACCESSION   CX098326
VERSION     CX098326.1  GI:56626295
KEYWORDS    EST.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Mann,B., Anderson,I. and Loftus,B.
  TITLE     Entamoeba histolytica EST reads
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Brendan Loftus
            The Institute for Genomic Research (TIGR; www.tigr.org)
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: bjloftus@tigr.org.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Entamoeba histolytica"
                     /mol_type="mRNA"
                     /strain="HM-1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="E. histolytica Normalized cDNA library"
ORIGIN

  Query Match             6.9%;  Score 46;  DB 8;  Length 617;
  Best Local Similarity   47.3%;  Pred. No. 0.63;
  Matches  139;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        262 GGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATA 321
              ||   |||  || | | ||   |  ||  ||       | | |  | || |   ||  ||
Db        235 GGACATGATACATTTATTAATTATAATTTAACAGTACTAGATGCTAATTACGTTAAAATT 294

Qy        322 GGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGG 381
              |||| |   |  | |||  |||   |||   | |   | |   |||||||  |
Db        295 GGTAATCATGTCTTGATTGGTCCTAATGTCCAACTCATTTCAGCTACTCATCCAACTGAT 354

Qy        382 TTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTT 441
                ||  |  |||           |   || |     |  |  |||   |||   || |
Db        355 CCACTTATTAGAAATTCTCTTGTTGAATATGGATTACCTATTGTAATTAAAGATGGTGCT 414

Qy        442 ACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATA 501
              |||  |  | ||  ||  |     |  |  |    ||   ||||  ||| ||| ||  | ||
Db        415 TGGATTGGTGCAGGTGCAACTATTCTTCCAGGAATTACAATTGGAGAAAATTCAGTTGTA 474

Qy        502 GGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTATTGGC 555
              ||  | | | |  ||| |||  || | ||||||  ||    ||  ||  ||  ||||
Db        475 GGTGCTGCTTCTGTAGTTACTCATGATGTTCCAGATAATACAGTTGTTGCTGGC 528


RESULT 7
CX099301
LOCUS       CX099301                622 bp    mRNA    linear   EST 15-DEC-2004
DEFINITION  EHAHY22TR E. histolytica Normalized cDNA library Entamoeba
            histolytica cDNA, mRNA sequence.
ACCESSION   CX099301
VERSION     CX099301.1  GI:56628282
KEYWORDS    EST.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Mann,B., Anderson,I. and Loftus,B.
  TITLE     Entamoeba histolytica EST reads
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Brendan Loftus
            The Institute for Genomic Research (TIGR; www.tigr.org)
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: bjloftus@tigr.org.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Entamoeba histolytica"
                     /mol_type="mRNA"
                     /strain="HM-1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="E. histolytica Normalized cDNA library"
ORIGIN

  Query Match             6.9%;  Score 46;  DB 8;  Length 622;
  Best Local Similarity   47.3%;  Pred. No. 0.63;
  Matches  139;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        262 GGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATA 321
              ||   |||  || | | ||   |  ||  ||       | | |  | || |   ||  ||
Db        234 GGACATGATACATTTATTAATTATAATTTAACAGTACTAGATGCTAATTACGTTAAAATT 293

Qy        322 GGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGG 381
              |||| |   |  | |||  |||   |||   | |   | |   |||||||  |
Db        294 GGTAATCATGTCTTGATTGGTCCTAATGTCCAACTCATTTCAGCTACTCATCCAACTGAT 353

Qy        382 TTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTT 441
                ||  |  |||           |   || |     |  |  |||   |||   || |
Db        354 CCACTTATTAGAAATTCTCTTGTTGAATATGGATTACCTATTGTAATTAAAGATGGTGCT 413

Qy        442 ACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATA 501
              |||  |  | ||  ||  |     |  |  |    ||   ||||  ||| ||| ||  | ||
Db        414 TGGATTGGTGCAGGTGCAACTATTCTTCCAGGAATTACAATTGGAGAAAATTCAGTTGTA 473
```

```
Qy        502 GGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTATTGGC 555
              ||  | | | | ||| |||  || | ||||||  ||    || || ||  ||||
Db        474 GGTGCTGCTTCTGTAGTTACTCATGATGTTCCAGATAATACAGTTGTTGCTGGC 527


RESULT 8
CX086360
LOCUS       CX086360                633 bp    mRNA    linear   EST 14-DEC-2004
DEFINITION  EHACI92TR E. histolytica Normalized cDNA library Entamoeba
            histolytica cDNA, mRNA sequence.
ACCESSION   CX086360
VERSION     CX086360.1  GI:56600354
KEYWORDS    EST.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 633)
  AUTHORS   Mann,B., Anderson,I. and Loftus,B.
  TITLE     Entamoeba histolytica EST reads
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Brendan Loftus
            The Institute for Genomic Research (TIGR; www.tigr.org)
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: bjloftus@tigr.org.
FEATURES             Location/Qualifiers
     source          1. .633
                     /organism="Entamoeba histolytica"
                     /mol_type="mRNA"
                     /strain="HM-1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="E. histolytica Normalized cDNA library"
ORIGIN

  Query Match             6.9%;  Score 46;  DB 8;  Length 633;
  Best Local Similarity   47.3%;  Pred. No. 0.64;
  Matches  139;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        262 GGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATA 321
              ||   |||  || | | ||   |  ||  ||      | | |  | || |   ||  ||
Db        234 GGACATGATACATTTATTAATTATAATTTAACAGTACTAGATGCTAATTACGTTAAAATT 293

Qy        322 GGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGG 381
              |||| |   |   | |||  |||   |||   | |   | |   |||||||  |
Db        294 GGTAATCATGTCTTGATTGGTCCTAATGTCCAACTCATTTCAGCTACTCATCCAACTGAT 353

Qy        382 TTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTT 441
              ||  |  |||         |   || |     |  |  |||   |||   || |
Db        354 CCACTTATTAGAAATTCTCTTGTTGAATATGGATTACCTATTGTAATTAAAGATGGTGCT 413

Qy        442 ACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATA 501
              |||  |  | ||  || |     |  |  |     ||   |||| ||| ||| ||  | ||
Db        414 TGGATTGGTGCAGGTGCAACTATTCTTCCAGGAATTACAATTGGAGAAAATTCAGTTGTA 473

Qy        502 GGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTATTGGC 555
              ||  | | | | ||| |||  || | ||||||  ||    || || ||  ||||
Db        474 GGTGCTGCTTCTGTAGTTACTCATGATGTTCCAGATAATACAGTTGTTGCTGGC 527


RESULT 9
CX093673
LOCUS       CX093673                633 bp    mRNA    linear   EST 15-DEC-2004
DEFINITION  EHAFO42TR E. histolytica Normalized cDNA library Entamoeba
            histolytica cDNA, mRNA sequence.
ACCESSION   CX093673
VERSION     CX093673.1  GI:56619485
KEYWORDS    EST.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 633)
  AUTHORS   Mann,B., Anderson,I. and Loftus,B.
  TITLE     Entamoeba histolytica EST reads
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Brendan Loftus
            The Institute for Genomic Research (TIGR; www.tigr.org)
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: bjloftus@tigr.org.
FEATURES             Location/Qualifiers
     source          1. .633
                     /organism="Entamoeba histolytica"
                     /mol_type="mRNA"
                     /strain="HM-1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="E. histolytica Normalized cDNA library"
ORIGIN

  Query Match             6.9%;  Score 46;  DB 8;  Length 633;
  Best Local Similarity   47.3%;  Pred. No. 0.64;
  Matches  139;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        262 GGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATA 321
              ||   |||  || | | ||   |  ||  ||      | | |  | || |   ||  ||
Db        234 GGACATGATACATTTATTAATTATAATTTAACAGTACTAGATGCTAATTACGTTAAAATT 293

Qy        322 GGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGG 381
              |||| |   |   | |||  |||   |||   | |   | |   |||||||  |
Db        294 GGTAATCATGTCTTGATTGGTCCTAATGTCCAACTCATTTCAGCTACTCATCCAACTGAT 353

Qy        382 TTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTT 441
              ||  |  |||         |   || |     |  |  |||   |||   || |
Db        354 CCACTTATTAGAAATTCTCTTGTTGAATATGGATTACCTATTGTAATTAAAGATGGTGCT 413

Qy        442 ACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATA 501
              |||  |  | ||  || |     |  |  |     ||   |||| ||| ||| ||  | ||
Db        414 TGGATTGGTGCAGGTGCAACTATTCTTCCAGGAATTACAATTGGAGAAAATTCAGTTGTA 473

Qy        502 GGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTATTGGC 555
              ||  | | | | ||| |||  || | ||||||  ||    || || ||  ||||
Db        474 GGTGCTGCTTCTGTAGTTACTCATGATGTTCCAGATAATACAGTTGTTGCTGGC 527


RESULT 10
CX083169
LOCUS       CX083169                635 bp    mRNA    linear   EST 14-DEC-2004
DEFINITION  EHAB786TR E. histolytica Normalized cDNA library Entamoeba
            histolytica cDNA, mRNA sequence.
ACCESSION   CX083169
VERSION     CX083169.1  GI:56597159
KEYWORDS    EST.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Mann,B., Anderson,I. and Loftus,B.
  TITLE     Entamoeba histolytica EST reads
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Brendan Loftus
            The Institute for Genomic Research (TIGR; www.tigr.org)
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: bjloftus@tigr.org.
FEATURES             Location/Qualifiers
     source          1. .635
                     /organism="Entamoeba histolytica"
                     /mol_type="mRNA"
                     /strain="HM-1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="E. histolytica Normalized cDNA library"
ORIGIN

  Query Match             6.9%;  Score 46;  DB 8;  Length 635;
```

```
  Best Local Similarity   47.3%;  Pred. No. 0.64;
  Matches  139;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        262 GGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATA 321
              ||   |||  || | | ||   |  ||  ||       | | |  | || |   ||  ||
Db        236 GGACATGATACATTTATTAATTATAATTTAACAGTACTAGATGCTAATTACGTTAAAATT 295

Qy        322 GGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGG 381
              ||||  |   |   | |||  |||   |||   | |   | |   |||||||  |
Db        296 GGTAATCATGTCTTGATTGGTCCTAATGTCCAACTCATTTCAGCTACTCATCCAACTGAT 355

Qy        382 TTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTT 441
              ||  |  |||         |   || |     |  |  |||   |||   || |
Db        356 CCACTTATTAGAAATTCTCTTGTTGAATATGGATTACCTATTGTAATTAAAGATGGTGCT 415

Qy        442 ACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATA 501
              |||  | | || || |   |  |     ||   |||| ||| ||| ||  | ||
Db        416 TGGATTGGTGCAGGTGCAACTATTCTTCCAGGAATTACAATTGGAGAAAATTCAGTTGTA 475

Qy        502 GGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTATTGGC 555
              ||  | | | | ||| |||  || | ||||||  ||    || || ||  ||||
Db        476 GGTGCTGCTTCTGTAGTTACTCATGATGTTCCAGATAATACAGTTGTTGCTGGC 529


RESULT 11
CX083304/c
LOCUS       CX083304                642 bp    mRNA    linear   EST 14-DEC-2004
DEFINITION  EHAB975TR E. histolytica Normalized cDNA library Entamoeba
            histolytica cDNA, mRNA sequence.
ACCESSION   CX083304
VERSION     CX083304.1  GI:56597294
KEYWORDS    EST.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   Mann,B., Anderson,I. and Loftus,B.
  TITLE     Entamoeba histolytica EST reads
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Brendan Loftus
            The Institute for Genomic Research (TIGR; www.tigr.org)
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: bjloftus@tigr.org.
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism="Entamoeba histolytica"
                     /mol_type="mRNA"
                     /strain="HM-1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="E. histolytica Normalized cDNA library"
ORIGIN

  Query Match             6.9%;  Score 46;  DB 8;  Length 642;
  Best Local Similarity   47.3%;  Pred. No. 0.64;
  Matches  139;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        262 GGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATA 321
              ||   |||  || | | ||   |  ||  ||       | | |  | || |   ||  ||
Db        400 GGACATGATACATTTATTAATTATAATTTAACAGTACTAGATGCTAATTACGTTAAAATT 341

Qy        322 GGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGG 381
              ||||  |   |   | |||  |||   |||   | |   | |   |||||||  |
Db        340 GGTAATCATGTCTTGATTGGTCCTAATGTCCAACTCATTTCAGCTACTCATCCAACTGAT 281

Qy        382 TTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTT 441
              ||  |  |||         |   || |     |  |  |||   |||   || |
Db        280 CCACTTATTAGAAATTCTCTTGTTGAATATGGATTACCTATTGTAATTAAAGATGGTGCT 221

Qy        442 ACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATA 501
              |||  | | || || |   |  |     ||   |||| ||| ||| ||  | ||
Db        220 TGGATTGGTGCAGGTGCAACTATTTTTCCAGGAATTACAATTGGAGAAAATTCAGTTGTA 161

Qy        502 GGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTATTGGC 555
              ||  | | | | ||| |||  || | ||||||  ||    || || ||  ||||
Db        160 GGTGCTGCTTCTGTAGTTACTCATGATGTTCCAGATAATACAGTTGTTGCTGGC 107


RESULT 12
CX091008
LOCUS       CX091008                663 bp    mRNA    linear   EST 15-DEC-2004
DEFINITION  EHAEM45TR E. histolytica Normalized cDNA library Entamoeba
            histolytica cDNA, mRNA sequence.
ACCESSION   CX091008
VERSION     CX091008.1  GI:56615376
KEYWORDS    EST.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   Mann,B., Anderson,I. and Loftus,B.
  TITLE     Entamoeba histolytica EST reads
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Brendan Loftus
            The Institute for Genomic Research (TIGR; www.tigr.org)
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: bjloftus@tigr.org.
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Entamoeba histolytica"
                     /mol_type="mRNA"
                     /strain="HM-1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="E. histolytica Normalized cDNA library"
ORIGIN

  Query Match             6.9%;  Score 46;  DB 8;  Length 663;
  Best Local Similarity   47.3%;  Pred. No. 0.64;
  Matches  139;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        262 GGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATA 321
              ||   |||  || | | ||   |  ||  ||       | | |  | || |   ||  ||
Db        268 GGACATGATACATTTATTAATTATAATTTAACAGTACTAGATGCTAATTACGTTAAAATT 327

Qy        322 GGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGG 381
              ||||  |   |   | |||  |||   |||   | |   | |   |||||||  |
Db        328 GGTAATCATGTCTTGATTGGTCCTAATGTCCAACTCATTTCAGCTACTCATCCAACTGAT 387

Qy        382 TTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTT 441
              ||  |  |||         |   || |     |  |  |||   |||   || |
Db        388 CCACTTATTAGAAATTCTCTTGTTGAATATGGATTACCTATTGTAATTAAAGATGGTGCT 447

Qy        442 ACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATA 501
              |||  | | || || |   |  |     ||   |||| ||| ||| ||  | ||
Db        448 TGGATTGGTGCAGGTGCAACTATTCTTCCAGGAATTACAATTGGAGAAAATTCAGTTGTA 507

Qy        502 GGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTATTGGC 555
              ||  | | | | ||| |||  || | ||||||  ||    || || ||  ||||
Db        508 GGTGCTGCTTCTGTAGTTACTCATGATGTTCCAGATAATACAGTTGTTGCTGGC 561


RESULT 13
AU037922
LOCUS       AU037922                600 bp    mRNA    linear   EST 29-MAR-1999
DEFINITION  AU037922 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSE457, mRNA sequence.
ACCESSION   AU037922
VERSION     AU037922.1  GI:3984675
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 600)
```

```
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSE457"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN

  Query Match             6.9%;  Score 45.8;  DB 1;  Length 600;
  Best Local Similarity   51.2%;  Pred. No. 0.71;
  Matches  107;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy          16 ATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAATAAAG 75
               |||  |    ||| ||    |   | | |    |||    |  |||| |||||| | ||
Db         344 ATTGGTCAAACTGCAACCGCTTCAACAAATGATATAAATTTATTAAAACAACAATTTAAA 403

Qy          76 CAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAAGAAA 135
               |||        |  | |   |||||  ||  |||| | |    |||| |  || |    |
Db         404 CAAATTCAAACTGTTAAAATTGCTCCAGTAGTTGTAGTTTCAATTTTTGGTGGTATCTTA 463

Qy         136 GTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAAG 195
               | | || |  |  |||  |  |||  ||  | ||||  ||| |  ||||   || ||
Db         464 TTAATTGCTGGTTTAGTAATGGCAATTAATGGTTTTCGTAAAACTTTTTATAATAATAAT 523

Qy         196 TATTTTAATGTTGCTATTGCTAATGATAA 224
               | | ||||| |  || |  || | || |
Db         524 CAATATAATGGTTATAATATTATTAATTA 552


RESULT 14
CNS000D1
LOCUS       CNS000D1               1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR01J16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065414
VERSION     AL065414.1  GI:4938827
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR01J16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             6.9%;  Score 45.8;  DB 10;  Length 1101;
  Best Local Similarity   14.8%;  Pred. No. 0.79;
  Matches   74;  Conservative  223;  Mismatches  201;  Indels    2;  Gaps    1;

Qy          29 GAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAATAAAGCAAGAATGTGGTT 88
               ||:| | ::::|::  || :|:|   |::   ::|  :    :   : : :|     ::|
Db         600 GARGATNDDKBAKDWKAAADAKAKAAAWHAGTWWTKKDTKKKKKKKKRKWGKAKRAATDDT 659

Qy          89 CTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAAGAAAGTTAATGGTTTTG 148
                 : | |:::   ::  :  : : ::|:|::  :  | :|:| |:|:|:: : ::::|:::
Db         660 AWTATTWKKTAKKDAGKAAKWKADRTKAWKGAKGAGARGKATAKARAKKRWKDDKDTKKK 719

Qy         149 AAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAAGTATTTTAATGTTG 208
               :|  |::       :    |:: ::: :     :::::|:: : | : :: ::||  |
Db         720 KTKTTATKKKTTTTKR--AATDKWKWRKKKATAGAWDWDDARRGRGGRWGDKRRRTGARG 777

Qy         209 CTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGGTTG 268
               :::|| :  ||:::::    |:  : ||| |: :::   :: ::: :: :::  ::: :::
Db         778 DKDKTGKKKKTGDKRDKGGGDGDRGDGGAGGGDRDRDARARRARWDTKDTRGAKDKADKK 837

Qy         269 AACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTG 328
               |   :::   |: :: |:  |:| :: :: :: :: :   :  :  : ||: : :::::
Db         838 KAGRKRKWWTRKTWKRDTADWKAWAAWRWRWRRDRKDRRGRRDRWRRWRGADKKWKWDDD 897

Qy         329 GCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTC 388
               :   :::| ::| :   | |:::  |::  :    ::: : :|::  |: |::  :|
Db         898 DTAKDDWTKKDTDWTAGATAGWDDDKAKDGDWKRGTDWRTWKRTWKGWGRKGDKGRKTAG 957

Qy         389 ATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACATTTG 448
               : |   :  : :: :  :: : |:    :::||:|  :: :::::  : ::   :::::
Db         958 RAGDGKDKKKGKDKDRTDDAKATDGKDTWRDTTDTAKRRRRKRRRRAGRGRDRKGDKKKR 1017

Qy         449 CTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATAGGCTCGG 508
                  :::|: :   ::::::|:::   : : :| :  ::| ||:  |:|:|:|:: :  :
Db        1018 WRARWRKGKKKKTKDKDDDDGKRWGRAKAKADAAKAKDDADAADAAAWAKAKAKDKDKAK 1077

Qy         509 GTGCAGTAATTAAGCAGGGT 528
               ::  |::: :: |: :::::
Db        1078 KKADADDRKDKKAKARKDDK 1097


RESULT 15
CL467398
LOCUS       CL467398               1295 bp    DNA     linear   GSS 31-MAR-2004
DEFINITION  SAIL_1269_F07.v2 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_1269_F07.v2, genomic survey sequence.
ACCESSION   CL467398
VERSION     CL467398.1  GI:45870303
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
```

```
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1295)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
   PUBMED   12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS847096; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .1295
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_1269_F07.v2"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN

  Query Match             6.9%;  Score 45.8;  DB 10;  Length 1295;
  Best Local Similarity   46.3%;  Pred. No. 0.81;
  Matches  184;  Conservative    0;  Mismatches  212;  Indels    1;  Gaps    1;

Qy          45 AACAATACCCATTCTAAATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGT 104
               || |  |   ||  || |  |  || |||| | |  ||    |    ||||| || |  |
Db         542 AATATAAAATATGATATAGTATTAACTAAAACTACTATTATATAAATACTATTCTATACT 601

Qy         105 TTTTGTGGATGATGTTTTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAA 164
               | ||  |  || | ||      ||   ||   ||   | | |  |   |  |  ||
Db         602 CTATGAATAACATATATTAAATATAATATAATAACATAAATATTATATATATATATACAT 661

Qy         165 CTGCTTTCTAAAAGCCCCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAA 224
                 ||  || |  |   |  | | || |  || |  |   |  |||    ||  || |
Db         662 AATATTAATATATACATATATATCATAAGATAATATACACTATAATATAATCAAAAATCA 721

Qy         225 GATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAA 284
               || || | | | | | | |   ||| | |  | |       | |   || || |||| |
Db         722 ATTATGAAATATATTAT-TTATATAATCTAAAAATAATTATATTATTGAAAAAATATATA 780

Qy         285 ACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCC 344
                 ||  |||||   |   |  | |  ||  |||   |  ||||      |||| || |
Db         781 TTATATAAATATAATAAATACTAAATATTATATATGATATAGTAATATAATTAATTATAA 840

Qy         345 CTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTC 404
               | | | | ||| || ||  | | |||||   |  | | |  ||    | |||||
Db         841 TAATATAATATTATATAATATAAATTATATATTAATATATATCAATTATCATATATAAAT 900

Qy         405 ATACGTTGCACATGATTGTCAAATAGGAGACTATGTT 441
               |||   |  | || ||  | | |||    ||  | ||
Db         901 ATAATATATAAATCATATTTATATAATTAACATTATT 937


Search completed: May 11, 2006, 05:00:37
Job time : 2137.07 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 20:34:42 ; Search time 92.8976 Seconds
(without alignments)
12743.667 Million cell updates/sec

Title: US-09-423-093B-2_COPY_13156_13821
Perfect score: 666
Sequence: 1 atgaatttgtatggtatttt.....gatcgccaacatctatttaa 666

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	666	100.0	14187	3	US-09-453-702B-121	Sequence 121, App
c	2	666	100.0	14187	3	US-10-114-170-121	Sequence 121, App
	3	218.2	32.8	687	3	US-09-328-352-752	Sequence 752, App
c	4	44.6	6.7	6876	3	US-08-956-171E-186	Sequence 186, App
c	5	44.6	6.7	6876	3	US-08-781-986A-186	Sequence 186, App
	6	44.4	6.7	657	3	US-09-543-681A-112	Sequence 112, App
	7	43.4	6.5	4605	3	US-09-543-681A-1010	Sequence 1010, Ap
c	8	43	6.5	720	3	US-09-710-279-1695	Sequence 1695, Ap
c	9	43	6.5	972	3	US-09-134-001C-2363	Sequence 2363, Ap
	10	43	6.5	3572	3	US-09-710-279-3988	Sequence 3988, Ap
	11	43	6.5	3806	3	US-09-710-279-4166	Sequence 4166, Ap
	12	41.8	6.3	2013	3	US-09-662-254B-53	Sequence 53, Appl
c	13	41.8	6.3	50000	3	US-09-662-254B-26	Sequence 26, Appl
c	14	41.6	6.2	1141	3	US-09-806-708B-22	Sequence 22, Appl
	15	41.6	6.2	1485	3	US-09-543-681A-601	Sequence 601, App
	16	41.2	6.2	822	3	US-09-328-352-3149	Sequence 3149, Ap
	17	41	6.2	67876	3	US-09-949-016-17232	Sequence 17232, A
	18	40.8	6.1	1664976	3	US-08-916-421B-1	Sequence 1, Appli
	19	40.8	6.1	1664976	3	US-09-692-570-1	Sequence 1, Appli
	20	40.6	6.1	601	3	US-09-949-016-120549	Sequence 120549,
c	21	40.6	6.1	133719	3	US-09-949-016-15092	Sequence 15092, A
	22	40	6.0	660	2	US-08-743-637B-175	Sequence 175, App
	23	40	6.0	660	3	US-08-526-840B-175	Sequence 175, App
	24	40	6.0	744	3	US-09-134-001C-2408	Sequence 2408, Ap
c	25	39.8	6.0	832	3	US-09-621-976-2813	Sequence 2813, Ap
	26	39.6	5.9	832	3	US-09-621-976-2813	Sequence 2813, Ap
	27	39.4	5.9	651	3	US-09-134-000C-2838	Sequence 2838, Ap
	28	39	5.9	1141	3	US-09-806-708B-22	Sequence 22, Appl
	29	38.8	5.8	519	3	US-09-134-000C-724	Sequence 724, App
	30	38.4	5.8	786431	3	US-09-751-389-3	Sequence 3, Appli
c	31	38	5.7	50000	3	US-09-662-254B-24	Sequence 24, Appl
c	32	37.8	5.7	1694	3	US-08-956-171E-339	Sequence 339, App
c	33	37.8	5.7	1694	3	US-08-781-986A-339	Sequence 339, App
c	34	37.8	5.7	7872	3	US-08-956-171E-161	Sequence 161, App
c	35	37.8	5.7	7872	3	US-08-781-986A-161	Sequence 161, App
	36	37.6	5.6	1830	3	US-09-662-254B-67	Sequence 67, Appl
	37	37.6	5.6	2304	3	US-09-662-254B-43	Sequence 43, Appl
c	38	37.6	5.6	32392	3	US-09-662-254B-27	Sequence 27, Appl
	39	37.6	5.6	50000	3	US-09-662-254B-24	Sequence 24, Appl
c	40	37.6	5.6	193303	3	US-09-497-855A-37	Sequence 37, Appl
c	41	37.6	5.6	193303	3	US-09-497-855A-44	Sequence 44, Appl
c	42	37.6	5.6	288031	3	US-09-949-016-14864	Sequence 14864, A
	43	37.4	5.6	2047	3	US-09-710-279-4461	Sequence 4461, Ap
	44	37.4	5.6	2169	3	US-09-710-279-1229	Sequence 1229, Ap
	45	37.4	5.6	2190	3	US-09-134-001C-1891	Sequence 1891, Ap

ALIGNMENTS

```
RESULT 1
US-09-453-702B-121/c
; Sequence 121, Application US/09453702B
; Patent No. 6365723
;   GENERAL INFORMATION:
;        APPLICANT: Blattner, Frederick R.
;                   Burland,     Valerie
;                   Perna,  Nicole T.
;                   Plunkett, Guy
;                   Welch,  Rod
;        TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
;        NUMBER OF SEQUENCES: 265
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Quarles & Brady
;             STREET: 1 South Pinckney Street
;             CITY: Madison
;             STATE: WI
;             COUNTRY: US
;             ZIP: 53701-2113
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Word Perfect 8.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/453,702B
;             FILING DATE: 03-Dec-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/110,955
;             FILING DATE: 04-DEC-1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seay, Nicholas J.
;             REGISTRATION NUMBER: 27386
;             REFERENCE/DOCKET NUMBER: 960296.95017
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (608) 251-5000
;             TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 121:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14187
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA  (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-09-453-702B-121
```

```
  Query Match           100.0%;  Score 666;  DB 3;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 4e-169;
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 786

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        785 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 726

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        725 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 666

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        665 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 606

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 546

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 486

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 426

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 366

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 306

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 246

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 186

Qy        661 ATTTAA 666
              ||||||
Db        185 ATTTAA 180


RESULT 2
US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Patent No. 6855814
;   GENERAL INFORMATION:
;        APPLICANT: Blattner, Frederick R.
;                   Burland,     Valerie
;                   Perna,  Nicole T.
;                   Plunkett, Guy
;                   Welch,  Rod
;        TITLE OF INVENTION: No. 6855814el Sequences of E. coli O157
;        NUMBER OF SEQUENCES: 265
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Quarles & Brady
;             STREET: 1 South Pinckney Street
;             CITY: Madison
;             STATE: WI
;             COUNTRY: US
;             ZIP: 53701-2113
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Word Perfect 8.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/114,170
;             FILING DATE: 01-Apr-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/453,702
;             FILING DATE: 03-DEC-1999
;             APPLICATION NUMBER: 60/110,955
;             FILING DATE: 04-DEC-1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seay, Nicholas J.
;             REGISTRATION NUMBER: 27386
;             REFERENCE/DOCKET NUMBER: 960296.95017
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (608) 251-5000
;             TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 121:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14187
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA  (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121

  Query Match           100.0%;  Score 666;  DB 3;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 4e-169;
  Matches  666;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 786

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        785 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 726

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        725 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 666

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        665 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 606

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 546

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 486

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 426

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 366

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 306

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        305 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 246

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 186

Qy        661 ATTTAA 666
              ||||||
Db        185 ATTTAA 180


RESULT 3
US-09-328-352-752
; Sequence 752, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 752
;   LENGTH: 687
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-752

  Query Match             32.8%;  Score 218.2;  DB 3;  Length 687;
  Best Local Similarity   60.1%;  Pred. No. 4.8e-49;
  Matches  383;  Conservative    0;  Mismatches  248;  Indels    6;  Gaps    1;

Qy         12 TGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAAT 71
              |||| ||| ||||||  |  ||| |||||  |||   || ||  |  |   ||| ||  |
Db         51 TGGTGTTTATGGTGCCAGTGGTTTTGGTAAGGAAGTTATGCCTTTAGTGCGTCAGCAGTT 110

Qy         72 AAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAA 131
                 |  | ||       | | |||      |||| | |||||||   |  | ||||
Db        111 TCCAACATTAAGTAAAGAACAATTTGC------TTTTATTGATGATGGACTATCTGGAAC 164

Qy        132 GAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAA 191
              ||   | ||||| |    ||| || ||      |  |||| | |     |||  |    |
Db        165 GACTTTAAATGGATACCCAGTACTATCCTATCTCGACTTTATCAGTAAACCTGCTGATCA 224

Qy        192 AAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAAT 251
              |||  | | |   | ||||| || |||  |  | | ||  |||   |   |       |
Db        225 TAAGGCTGTGACAATCGCTATAGCAAATAGTGTGGTTCGTGAGAAGTTAGTTAGCCTTTT 284

Qy        252 ATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCA 311
              |  |  | | || ||  |||   || || |  ||  | ||||||  ||  ||  |||| |
Db        285 AGAAAAAGATGGTGTACAACACCTAGCTGTTCAATCTACAAATACAGTAATTCTTGATGA 344

Qy        312 TACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCA 371
                |  |||||||   ||     |  ||| |||||||  |    || || ||||||   |
Db        345 GGTTGAGATAGGTGAAGGAAGCTTACTTTGTCCCTTTACTTGCTTAACTTCTAATATAAA 404

Qy        372 TATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGG 431
              || || | |||||| || || || ||    || || || |||||||||  ||||||
Db        405 AATTGGAAAGTTTTTCCACGCTAATATTTATAGTTATGTGGCTCATGATTGTGTAATAGG 464

Qy        432 AGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAA 491
              ||| |||||||||||| || ||||| ||||| ||||| ||  || ||  |||||||||  |
Db        465 AGATTATGTTACATTCGCGCCTGGCGCTAAGTGTAACGGTAATATTCATATTGAAGATCA 524

Qy        492 TGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTAT 551
              ||| || || ||| | || ||||| ||||| |||||    ||| ||   ||  | || ||
Db        525 TGCCTACATTGGCACTGGAGCAGTTATTAAACAGGGCACACCTGATAAGCCTTTAATAAT 584

Qy        552 TGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTTCCTGCCGGTAT 611
              |||    || || ||| | || ||||| ||||| || || |||||||| || || ||  |
Db        585 TGGTAAAGGTGCTATTGTTGGAATGGGCGCTGTAGTAACGAAAAGTGTGCCAGCTGGCGT 644

Qy        612 AACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGA 648
              ||||||    || ||||| ||||||    |  |||||
Db        645 AACTGTTGTAGGTAATCCTGCAAGAATTTTAGAAAGA 681


RESULT 4
US-08-956-171E-186/c
; Sequence 186, Application US/08956171E
; Patent No. 6593114
;   GENERAL INFORMATION:
;        APPLICANT: Charles Kunsch
;                   Gil H. Choi
;                   Patrick S. Dillon
;                   Craig A. Rosen
;                   Steven C. Barash
;                   Michael R. Fannon
;        TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;        NUMBER OF SEQUENCES: 5256
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/956,171E
;             FILING DATE: 20-Oct-1997
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/009,861
;             FILING DATE: January 5, 1996
;             APPLICATION NUMBER: 08/781,986
;             FILING DATE: January 3, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mark J. Hyman
;             REGISTRATION NUMBER: 46,789
;             REFERENCE/DOCKET NUMBER: PB248P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (240) 314-1224
;             TELEFAX: (301) 309-8439
;   INFORMATION FOR SEQ ID NO: 186:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 6876 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 186:
US-08-956-171E-186

  Query Match              6.7%;  Score 44.6;  DB 3;  Length 6876;
  Best Local Similarity   54.6%;  Pred. No. 0.052;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              | |||||||||  |||||  | | ||  | || || |||  |   |||  || ||   ||
Db       1598 ATATAGGAGACAATGTTTTAATAGCAGCCGGTGCGAAAGTTTTAGGAAATATTAAAATAA 1539

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |    ||||  | ||||  ||  |   | ||||  ||   ||    |||||  | |    ||
Db       1538 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 1479

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587
                 || ||||       ||    ||| | || |   | || |||
Db       1478 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 1436


RESULT 5
```

```
US-08-781-986A-186/c
; Sequence 186, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
;   APPLICANT: Charles Kunsch
;   TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;   NUMBER OF SEQUENCES: 5255
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Human Genome Sciences, Inc.
;     STREET: 9410 Key West Avenue
;     CITY: Rockville
;     STATE: Maryland
;     COUNTRY: USA
;     ZIP: 20850
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;     COMPUTER: HP Vectra 486/33
;     OPERATING SYSTEM: MSDOS version 6.2
;     SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/781,986A
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Benson, Bob
;     REGISTRATION NUMBER: 30,446
;     REFERENCE/DOCKET NUMBER: PB248PP
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (301) 309-8504
;     TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 186:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 6876 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
US-08-781-986A-186

  Query Match             6.7%;  Score 44.6;  DB 3;  Length 6876;
  Best Local Similarity   54.6%;  Pred. No. 0.052;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              | ||||||||| |||||  | | ||  | || || |||  |  ||| || ||   ||
Db       1598 ATATAGGAGACAATGTTTTAATAGCAGCCGGTGCGAAAGTTTTAGGAAATATTAAAATAA 1539

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |   |||| | ||||| ||  |   | ||||  ||   ||     ||||| | |     ||
Db       1538 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 1479

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587
                || ||||       ||    ||| | || |    | || |||
Db       1478 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 1436


RESULT 6
US-09-543-681A-112
; Sequence 112, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 112
;   LENGTH: 657
;   TYPE: DNA
;   ORGANISM: Proteus mirabilis
US-09-543-681A-112

  Query Match             6.7%;  Score 44.4;  DB 3;  Length 657;
  Best Local Similarity   55.1%;  Pred. No. 0.028;
  Matches   87;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        459 TAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAAT 518
              ||   |||| ||  ||  |  |||| | ||   || ||  |||||    | | ||| | ||
Db        330 TATTGGTAAAGGTGATACTTATATTCACGATGGTGTATGGATAGGTATGCGGGCAATGAT 389

Qy        519 TAAGCAGGGTGTTCCTAATCGCCCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGG 578
              || ||  || ||| ||| | |   |      |||| ||| | |   | | |||  |
Db        390 TATGCCTGGCGTTACTATTGGTGAAGGGGCTATTATCGCAGCAAATAGTGTAGTGACAAA 449

Qy        579 GGCTGTTGTCACTAAAAGTGTTCCTGCCGGTATAACTG 616
                | |||||   || | ||| |   ||  || ||| | |
Db        450 AGATGTTGCTCCTTATAGTATCGTTGGTGGAATACCGG 487


RESULT 7
US-09-543-681A-1010
; Sequence 1010, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 1010
;   LENGTH: 4605
;   TYPE: DNA
;   ORGANISM: Proteus mirabilis
US-09-543-681A-1010

  Query Match             6.5%;  Score 43.4;  DB 3;  Length 4605;
  Best Local Similarity   42.2%;  Pred. No. 0.097;
  Matches  242;  Conservative    0;  Mismatches  331;  Indels    0;  Gaps    0;

Qy         11 ATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAA 70
               ||   ||| || |  |    | | | ||  ||| | ||  |  | || | || |||
Db        356 ATTTGATTATTAGAACCAAGAAATTGGTTAATGAATCTATTACTCCTTTATTTAAAGAAA 415

Qy         71 TAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAA 130
              ||||||||||    || ||||    |  || |   | | |    | || || |
Db        416 TAAAGCAAGAGCCTGATTCTACATCGTCTATAAATATAGGTACTCATTTTCTTCATCCTT 475

Qy        131 AGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAA 190
              | ||   |  |      | |  | |||  ||  |  |    ||    ||        | |
Db        476 ACAATAGTCCTACGAATAATTTCCTTATAATTATTTTACTTGCTGAATTAAAAGAACTCT 535

Qy        191 AAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAA 250
                |  |        ||| |||      ||   || |         | ||   ||    |
Db        536 CTGATAAAAAGTGGTTTGATATATTAGATTTAAAAAATAATGGAGGTGTTAATGTAGAAC 595

Qy        251 TATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATC 310
                | ||||  |    |   ||   | |      |  | |     || | |  ||
Db        596 GCTCATTATTTGAAAAGAATAAAAAGTATGTTCCTAGTGTTATTTATTCTGGTAAAAATT 655

Qy        311 ATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTC 370
              |||   |  ||  |    | ||| | | | | || ||   |  | ||||||  ||
Db        656 TATATATACAATCTAACGAGTTTTTTAATACCCAATCTGATATTTTTGCAACTAATGATC 715

Qy        371 ATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAG 430
              |||  |  |        ||||||| ||||||  ||            |  | | | |
Db        716 TTATCGCAACGGGAAACAATGCAAAAGTATACTATTATCGGTTTGGACACCTTAATAAAT 775
```

```
Qy        431 GAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACA 490
              | ||   |  | |   || |  ||    ||  | | || || ||| ||| | | || | |
Db        776 GGGATCTTTATTCTCATGATAATGCTCATATTTTTTATCGAGATGATGTAAATAAATATA 835

Qy        491 ATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTA 550
              ||    ||  |      |  |  |||    ||   | ||||   |     | ||   |
Db        836 ATATTGATGAAACTAAATTTCAAAAAATAGTTCATTTTATTCCCTTTACTAAAATTGGAA 895

Qy        551 TTGGCGCGGGAGCCATTATAGGTATGGGGGCTG 583
              |      | |   |||  |||  ||  |   |||
Db        896 GTCATTATGAATTTATTGTAGATAGTGCTTCTG 928


RESULT 8
US-09-710-279-1695/c
; Sequence 1695, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1695
;   LENGTH: 720
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-1695

  Query Match             6.5%;  Score 43;  DB 3;  Length 720;
  Best Local Similarity   50.2%;  Pred. No. 0.068;
  Matches  106;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy         15 TATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAATAAA 74
              |||| |   || |   |   ||| | |||  | |||    ||   | | ||  ||||||
Db        618 TATTGTACCTGATATCAAAGATGCTGGAGCGATAATCATGATAGGATGTGAAGTAATAAA 559

Qy         75 GCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAAGAA 134
               |      || || | ||  ||| |      | ||  |||| || ||  |||   ||
Db        558 TTCACGCCCTGCTTGTAACATTGCCCCCCACTCTGGAGATGGTGGTTGTGCACCTAAACC 499

Qy        135 AGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAA 194
                |||| |  |||| |  | |  |  |      |||| |||| |  ||    ||| ||
Db        498 TAAAAATGATAATGAACTAATGTATAAAATGATTTTTCCAAAATCAACTATCATAATAAC 439

Qy        195 GTATTTTAATGTTGCTATTGCTAATGATAAG 225
              || |  ||| |   | |  |  | |||||
Db        438 AAATATAGATGGTAAAACTTTTGGTAATAAG 408


RESULT 9
US-09-134-001C-2363/c
; Sequence 2363, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2363
;   LENGTH: 972
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2363

  Query Match             6.5%;  Score 43;  DB 3;  Length 972;
  Best Local Similarity   50.2%;  Pred. No. 0.075;
  Matches  106;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy         15 TATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAATAAA 74
              |||| |   || |   |   ||| | |||  | |||    ||   | | ||  ||||||
Db        870 TATTGTACCTGATATCAAAGATGCTGGAGCGATAATCATGATAGGATGTGAAGTAATAAA 811

Qy         75 GCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAAGAA 134
               |      || || | ||  ||| |      | ||  |||| || ||  |||   ||
Db        810 TTCACGCCCTGCTTGTAACATTGCCCCCCACTCTGGAGATGGTGGTTGTGCACCTAAACC 751

Qy        135 AGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAA 194
                |||| |  |||| |  | |  |  |      |||| |||| |  ||    ||| ||
Db        750 TAAAAATGATAATGAACTAATGTATAAAATGATTTTTCCAAAATCAACTATCATAATAAC 691

Qy        195 GTATTTTAATGTTGCTATTGCTAATGATAAG 225
              || |  ||| |   | |  |  | |||||
Db        690 AAATATAGATGGTAAAACTTTTGGTAATAAG 660


RESULT 10
US-09-710-279-3988
; Sequence 3988, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3988
;   LENGTH: 3572
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3988

  Query Match             6.5%;  Score 43;  DB 3;  Length 3572;
  Best Local Similarity   50.2%;  Pred. No. 0.11;
  Matches  106;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy         15 TATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAATAAA 74
              |||| |   || |   |   ||| | |||  | |||    ||   | | ||  ||||||
Db       2178 TATTGTACCTGATATCAAAGATGCTGGAGCGATAATCATGATAGGATGTGAAGTAATAAA 2237

Qy         75 GCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAAGAA 134
               |      || || | ||  ||| |      | ||  |||| || ||  |||   ||
Db       2238 TTCACGCCCTGCTTGTAACATTGCCCCCCACTCTGGAGATGGTGGTTGTGCACCTAAACC 2297

Qy        135 AGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAA 194
                |||| |  |||| |  | |  |  |      |||| |||| |  ||    ||| ||
Db       2298 TAAAAATGATAATGAACTAATGTATAAAATGATTTTTCCAAAATCAACTATCATAATAAC 2357

Qy        195 GTATTTTAATGTTGCTATTGCTAATGATAAG 225
              || |  ||| |   | |  |  | |||||
Db       2358 AAATATAGATGGTAAAACTTTTGGTAATAAG 2388


RESULT 11
```

```
US-09-710-279-4166
; Sequence 4166, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4166
;   LENGTH: 3806
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-4166

  Query Match             6.5%;  Score 43;  DB 3;  Length 3806;
  Best Local Similarity   50.2%;  Pred. No. 0.12;
  Matches  106;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy         15 TATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAATAAA 74
              |||| |   || |   |   ||| | |||  | |||    ||   |  | ||  ||||||
Db        133 TATTGTACCTGATATCAAAGATGCTGGAGCGATAATCATGATAGGATGTGAAGTAATAAA 192

Qy         75 GCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAAGAA 134
                |    || || | ||  ||| |      | ||  |||| || ||   |||    ||
Db        193 TTCACGCCCTGCTTGTAACATTGCCCCCCACTCTGGAGATGGTGGTTGTGCACCTAAACC 252

Qy        135 AGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAA 194
                |||| |  |||| |  | | |  |  |     |||| |||| |  ||    ||| ||
Db        253 TAAAAATGATAATGAACTAATGTATAAAATGATTTTTCCAAAATCAACTATCATAATAAC 312

Qy        195 GTATTTTAATGTTGCTATTGCTAATGATAAG 225
                || |  ||| |   | | |  | |||||
Db        313 AAATATAGATGGTAAAACTTTTGGTAATAAG 343


RESULT 12
US-09-662-254B-53
; Sequence 53, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
;   LENGTH: 2013
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (1)..(2013)
;   OTHER INFORMATION:
US-09-662-254B-53

  Query Match             6.3%;  Score 41.8;  DB 3;  Length 2013;
  Best Local Similarity   51.3%;  Pred. No. 0.2;
  Matches   97;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        187 TTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAG 246
              ||| |  | |     |||| |  |||| | ||  |||      | |           | ||
Db         16 TTAGATGATTTGAGAAATGATAATATTCCGAACAATATTGCTCAAATTTTGCCTCATCAG 75

Qy        247 TCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTAT 306
              | |  |  |||| |     |  | | || |  ||||||  || | ||||| ||| | |
Db         76 TTAGCAACATTAGATTTTTTATATCAAAAATGTATAAATAATGCTAATAGTGTTTTATTG 135

Qy        307 GATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAAT 366
              |||||  |||     | ||    | ||||||||||  ||  |||  ||||  |    |
Db        136 TTTCATAAAATGGGTTCTGGTAAAACAATTATTTCTCTTTTGTTTAGCATATTAATATGT 195

Qy        367 ACTCATATA 375
              | |  || |
Db        196 AATATTAAA 204


RESULT 13
US-09-662-254B-26/c
; Sequence 26, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-26

  Query Match             6.3%;  Score 41.8;  DB 3;  Length 50000;
  Best Local Similarity   51.3%;  Pred. No. 0.56;
  Matches   97;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        187 TTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAG 246
              ||| |  | |     |||| |  |||| | ||  |||      | |           | ||
Db       2710 TTAGATGATTTGAGAAATGATAATATTCCGAACAATATTGCTCAAATTTTGCCTCATCAG 2651

Qy        247 TCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTAT 306
              | |  |  |||| |     |  | | || |  ||||||  || | ||||| ||| | |
Db       2650 TTAGCAACATTAGATTTTTTATATCAAAAATGTATAAATAATGCTAATAGTGTTTTATTG 2591

Qy        307 GATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAAT 366
              |||||  |||     | ||    | ||||||||||  ||  |||  ||||  |    |
Db       2590 TTTCATAAAATGGGTTCTGGTAAAACAATTATTTCTCTTTTGTTTAGCATATTAATATGT 2531

Qy        367 ACTCATATA 375
              | |  || |
Db       2530 AATATTAAA 2522


RESULT 14
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
```

```
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             6.2%;  Score 41.6;  DB 3;  Length 1141;
  Best Local Similarity   8.5%;  Pred. No. 0.19;
  Matches   45;  Conservative  236;  Mismatches  250;  Indels    1;  Gaps    1;

Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              |:::::::  : :  : |:|   | |:: |:::: ::  | : :: :|: ::::| : |:
Db        599 ADSWVWWYANWMRCRDVTYTRNNTYCKSYAHSYWYWSNNAMWYRRYSARNWSSMARWTTR 540

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
               : :   :: ::|| :: :::    |: |: ::::: :  : || :::   |:::::
Db        539 NNWWMSGBVRMRWAGTMWWRHWNNNNTD-TRYYWWWKRWARBTTTVYDSMCNAKSMWRGN 481

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              :: :   :|| :|| ::: :     :  :  :::: : :     :       :| :
Db        480 NWRAMKMWWAANNDAGAMDHWTYWMGNNTMWMRRAWKMMNMAWCRRAYCCNNNNNRACVW 421

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              : :::: : :  ::: |      :  :   |:  : ::|  |:: :: ::::: :  ::
Db        420 HKHKMWRWTWKYMWKAACNNNNBKAMYMRVAWMMYSRDTTNTDWMMWTSDWBWHWYTVDY 361

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              |: :|        :::    :     : :  :  :    |:  |  :| ::::|   : :
Db        360 TMMRAWNNNNNNNWRBCKTTSWMWWMDHMNTHCTYGNNTWGSAYBMAAMSMWAAGASNBV 301

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              : : :::|::          : : ::   : :    ::::  ::|::      :|:::
Db        300 TYNWCWRMTYMGKTMTNNNNNNKAWYYRTKTVAWCNNRYYYDTAVWTBKRNYKYCYAYBW 241

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              :::    :::::::: :::::  ::       ::::: :|: : :: :  ||   :::
Db        240 YYBMYMGKHHWBWWRRABHRSWNMWWVKCRNKYMVSWHYHAMRYBKWABAVGCNNNWKDR 181

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ::::       : ::: : :::::|    |  :  :::::: :::  :::  ::
Db        180 MAHHHWCATNNNMMWWWYAYMHHMHKKGKAAWTNNKTABRDDHBAHVKTYWYWRYDYWCA 121

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTC 532
              : :  |:| ::| : ::: : ::: :   ||::  :: : :   :   ::::
Db        120 MCWMNAKAKVRTAMKHMWYYTDRYVSANNTGVRWMMMRWCMWWYSMNNRWYY 69


RESULT 15
US-09-543-681A-601
; Sequence 601, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;  APPLICANT: GARY BRETON
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;  TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.1002-001
;  CURRENT APPLICATION NUMBER: US/09/543,681A
;  CURRENT FILING DATE:  2000-04-05
;  PRIOR APPLICATION NUMBER: US 60/128,706
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 601
;   LENGTH: 1485
;   TYPE: DNA
;   ORGANISM: Proteus mirabilis
US-09-543-681A-601

  Query Match             6.2%;  Score 41.6;  DB 3;  Length 1485;
  Best Local Similarity   46.1%;  Pred. No. 0.21;
  Matches  140;  Conservative    0;  Mismatches  164;  Indels    0;  Gaps    0;

Qy        292 AATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTT 351
              |||    | || | || | || |||    |       |   || |||||         |||
Db        172 AATGTAATGGTATTTGGTAATGCTAAAGTCACAGATAATGCAATTATAAGAAATAATGTG 231

Qy        352 ACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTT 411
              |  || |    |||| |     |   | |   |  | | | |||     |  || |  ||
Db        232 AAGATCTACGGTAATGCAATAGTTAAGGGAAATAGTAAAGTAAAAGACAATGCAGAGATT 291

Qy        412 GCACATGATTGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGA 471
                | ||   | | | | |||  |||| ||| |  ||  ||  | |  |    |||
Db        292 TATGGTAATGTGCTAGTGGAAGATAATGTGACAATATCTGATGATGTTGTAATATATGAT 351

Qy        472 TATGTTGTTATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTT 531
              |||  |||||| |||| |||||    |||       | ||| |||||  |  |   || |
Db        352 AATGCAGTTATTAAAGATAATGCGAGAATATCTGACGATGCTGTAATATATGATAATGCT 411

Qy        532 CCTAATCGCCCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACT 591
              || |        |  ||  |   | | |     ||| | |   || | ||   |   |
Db        412 GTTATTAAAGATAATGCTAAAGTATCAGAATATGCTATTGTTCGTGGTGATGCAATAGTT 471

Qy        592 AAAA 595
              |||
Db        472 GAAA 475


Search completed: May 11, 2006, 05:23:01
Job time : 96.8976 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:01:11 ; Search time 439.58 Seconds
(without alignments)
12528.801 Million cell updates/sec

Title: US-09-423-093B-2_COPY_13156_13821
Perfect score: 666
Sequence: 1 atgaatttgtatggtatttt.....gatcgccaacatctatttaa 666

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	666	100.0	14024	3	US-09-423-093A-2	Sequence 2, Appli
c 2	666	100.0	14187	5	US-10-114-170-121	Sequence 121, App
3	71.4	10.7	1199	7	US-10-282-122A-35663	Sequence 35663, A
4	46.6	7.0	1463	7	US-10-437-963-81558	Sequence 81558, A
5	45.2	6.8	645	3	US-09-738-626-406	Sequence 406, App
6	45.2	6.8	3309400	3	US-09-738-626-1	Sequence 1, Appli
7	44.6	6.7	627	3	US-09-815-242-4528	Sequence 4528, Ap
8	44.6	6.7	642	7	US-10-282-122A-8128	Sequence 8128, Ap
9	44.6	6.7	681	3	US-09-815-242-8357	Sequence 8357, Ap
10	44.6	6.7	681	3	US-09-815-242-8685	Sequence 8685, Ap
11	44.6	6.7	3700	8	US-10-857-625-15	Sequence 15, Appl
12	44.6	6.7	4350	8	US-10-857-625-14	Sequence 14, Appl
c 13	44.6	6.7	6876	2	US-08-781-986A-186	Sequence 186, App
c 14	44.6	6.7	6876	7	US-10-329-624-186	Sequence 186, App
c 15	43	6.5	972	7	US-10-724-972A-460	Sequence 460, App
16	42	6.3	720	7	US-10-282-122A-35000	Sequence 35000, A
17	41.8	6.3	2013	8	US-10-706-635-53	Sequence 53, Appl
c 18	41.8	6.3	50000	8	US-10-706-635-26	Sequence 26, Appl
19	41.8	6.3	1163020	7	US-10-398-221-10	Sequence 10, Appl
20	41.8	6.3	3011208	7	US-10-398-221-2058	Sequence 2058, Ap
21	41.6	6.2	660	7	US-10-282-122A-24994	Sequence 24994, A
c 22	41.6	6.2	1500	9	US-10-955-924B-7	Sequence 7, Appli
23	41.4	6.2	3411	5	US-10-027-632-262354	Sequence 262354,
24	41.4	6.2	3411	6	US-10-027-632-262354	Sequence 262354,
c 25	41.4	6.2	3422	7	US-10-250-987-1	Sequence 1, Appli
26	41.4	6.2	3673778	6	US-10-312-841-2	Sequence 2, Appli
c 27	41.2	6.2	1200	5	US-10-027-632-117254	Sequence 117254,
c 28	41.2	6.2	1200	6	US-10-027-632-117254	Sequence 117254,
29	41	6.2	6763	7	US-10-433-793-98	Sequence 98, Appl
30	41	6.2	29607	3	US-09-764-877-3626	Sequence 3626, Ap
31	41	6.2	29607	6	US-10-242-515-3626	Sequence 3626, Ap
32	40.8	6.1	711	7	US-10-282-122A-24586	Sequence 24586, A
c 33	40.8	6.1	1119	7	US-10-250-987-2	Sequence 2, Appli
c 34	40.8	6.1	1911	8	US-10-723-860-8112	Sequence 8112, Ap
c 35	40.8	6.1	2873	5	US-10-116-802-418	Sequence 418, App
36	40.8	6.1	2939	3	US-09-764-868-1314	Sequence 1314, Ap
c 37	40.8	6.1	2939	3	US-09-764-868-1315	Sequence 1315, Ap
38	40.8	6.1	40324	7	US-10-433-793-180	Sequence 180, App
39	40.2	6.0	1310	4	US-09-925-065A-705471	Sequence 705471,
40	40	6.0	633	7	US-10-282-122A-15630	Sequence 15630, A
41	40	6.0	660	3	US-09-452-599-175	Sequence 175, App
42	40	6.0	660	6	US-10-121-120-175	Sequence 175, App
43	40	6.0	660	8	US-10-121-120-175	Sequence 175, App
44	40	6.0	744	7	US-10-724-972A-614	Sequence 614, App
45	40	6.0	3673778	6	US-10-312-841-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-423-093A-2
; Sequence 2, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 14024
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-423-093A-2

Query Match 100.0%; Score 666; DB 3; Length 14024;
Best Local Similarity 100.0%; Pred. No. 1.8e-154;
Matches 666; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13156 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 13215

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13216 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 13275

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13276 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 13335

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13336 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 13395

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13396 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 13455
```

```
Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13456 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 13515

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13516 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 13575

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13576 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 13635

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13636 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 13695

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13696 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 13755

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13756 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 13815

Qy        661 ATTTAA 666
              ||||||
Db      13816 ATTTAA 13821
```

RESULT 2
US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14187
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121

Query Match 100.0%; Score 666; DB 5; Length 14187;
Best Local Similarity 100.0%; Pred. No. 1.8e-154;
Matches 666; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 ATGAATTTGTATGGTATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTA 786

Qy         61 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        785 AATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTT 726

Qy        121 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        725 TTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCC 666

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        665 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 606

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 546

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 486

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGAT 426

Qy        421 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 TGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTT 366

Qy        481 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 ATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGC 306

Qy        541 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 CCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTT 246

Qy        601 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 CCTGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGATCGCCAACATCT 186

Qy        661 ATTTAA 666
              ||||||
Db        185 ATTTAA 180
```

RESULT 3
US-10-282-122A-35663
; Sequence 35663, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John

```
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35663
;   LENGTH: 1199
;   TYPE: DNA
;   ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-35663

  Query Match             10.7%;  Score 71.4;  DB 7;  Length 1199;
  Best Local Similarity   48.6%;  Pred. No. 3.3e-07;
  Matches  195;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy        194 AGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAATAT 253
              | |  |||| | |||||||||  | | || | || || || | | | | | |   | ||
Db        769 ATTCATTTATTATTGCTATTGGAAGTAATCATATTCGTCAAAAACTTTTTCATAAATTAG 828

Qy        254 TATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATA 313
              | |  |    | | ||     |||  |||  |||||||     |||||   |  |  |
Db        829 AAGTTAAAATTGATCAATATGCAACATTAATTCATCCAACAGCAGTTGTGAGTCCTAGTG 888

Qy        314 CTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATA 373
              | | ||| |||   ||  |  || |    ||   ||     |||  | |  |  ||   |
Db        889 CGAAGATTGGTTACGGAACGGTTGTAATGCCTAATGCGGTGATAAATGCGGACTCTATAA 948

Qy        374 TAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAG 433
              | || | |  | |       |||| |  | ||   || | ||| |||  ||||| || |
Db        949 TTGGAAAGCATGTCATAATCAACACAGGCGCAATAGTCGAACACGATAATCAAATTGGTG 1008

Qy        434 ACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATG 493
              | ||||||    || | |||   |||| ||     ||   ||||    | |  ||
Db       1009 ATTATGTTCATATTTCACCTAATGCTACATTAGCAGGTGGTGTTAAAGTAGGCGAGGCAT 1068

Qy        494 CATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCACTTATTATTG 553
              || | || |       ||| |   | |  | |    ||   | | |   |     |||||
Db       1069 CACACATTGCAATTAATGCTGCTGTAATACCGTTGATTGAAATTGGTGAAAGATGTATTG 1128

Qy        554 GCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAA 594
                |  | ||     | ||  ||      ||||   ||  ||
Db       1129 TGGGAGCAGGTGCAACAGTCATAAATAATGTTAAAACAGAA 1169


RESULT 4
US-10-437-963-81558
; Sequence 81558, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 81558
;   LENGTH: 1463
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_81072C.1
US-10-437-963-81558

  Query Match             7.0%;  Score 46.6;  DB 7;  Length 1463;
  Best Local Similarity   48.3%;  Pred. No. 0.53;
  Matches  130;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        365 ATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCATACGTTGCACATGATTGTC 424
              |||    ||| ||       |  || |  ||      |  |   |  | ||||||||||
Db        522 ATAAGGTTATTGGTGACAATAATCTTATAATGGGTTCATGCCATATAGCACATGATTGCA 581

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              || ||  || ||   | |||||||        ||  || |  |||  ||||||| |||
Db        582 GGATTGGCAACAATAACATATTTGCTAACAATACTCTATTTGCTGGCCATGTTGTTGTTG 641

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              ||||     |  |||  |    || ||  ||  |  | ||   |  || || |  ||
Db        642 AAGATTGCACCCATACTGCAGGGGCTGTTGTTGTCCATCAATTTTGTCATATTGGGTCAT 701

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTTCCTG 604
              ||    ||   |  |||| |  | |    |    | |    || |  ||  ||| ||
Db        702 TTTCATTTCTAGGTGGAGGCTCTGTGTTCAATTTGAATCAGATCGCACAAGATGTGCCAA 761

Qy        605 CCGGTATAACTGTGTGCGGAAATCCAGCA 633
                  || |  ||    ||  ||  ||||
Db        762 GATACATGATGGTTGCAGGTGATAGAGCA 790


RESULT 5
US-09-738-626-406
; Sequence 406, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
```

```
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 406
;    LENGTH: 645
;    TYPE: DNA
;    ORGANISM: Corynebacterium glutamicum
US-09-738-626-406

  Query Match             6.8%;  Score 45.2;  DB 3;  Length 645;
  Best Local Similarity   46.1%;  Pred. No. 0.81;
  Matches  152;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy        206 TTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGG 265
              |||| ||||  ||   | ||| || || |   | ||||      ||         ||
Db        230 TTGCGATTGGAAACCCAACGATTCGCCATAAGCTTTCTGCTCTTTTAGACAAGTCTGGTT 289

Qy        266 TTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTA 325
              | ||  || | ||       |||||    |   ||  |        ||      || | ||
Db        290 TCGATGCATTTACGGCGGTGCATCCCTCTGCAGTGATAGGAAGCCAAGTAGTGTTTGGGT 349

Qy        326 GTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTT 385
              ||| ||| || | | |     |      || || || |||       | ||   | | |
Db        350 CTGGGGCTGTTGTATGTTCTGGGGCCGTGATTTCGACCAATGTAAGGCTTGGTGTGCATG 409

Qy        386 TTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACAT 445
              | |||   |||  | |         | |  |||||||       | | ||||||||| |
Db        410 TGCATATCAACCCAAATGTGACAATCGGGCATGATTCAATTCTTGAAGACTATGTTTCGA 469

Qy        446 TTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATAGGCT 505
              |   ||| |  || | |  |   ||  | ||  | ||    ||  |  |   ||| ||
Db        470 TAAATCCGGCTGCAACAATTTCGGGTGAAGTCCTGATCATGGAAGAAACTCTTATTGGAG 529

Qy        506 CGGGTGCAGTAATTAAGCAGGGTGTTCCTA 535
              | |  ||   | |   ||| ||  |  |||
Db        530 CTGCGGCTACAGTATTGCAAGGACTGACTA 559


RESULT 6
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
;    LENGTH: 3309400
;    TYPE: DNA
;    ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

  Query Match             6.8%;  Score 45.2;  DB 3;  Length 3309400;
  Best Local Similarity   46.1%;  Pred. No. 43;

  Matches  152;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy        206 TTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGG 265
              |||| ||||  ||   | ||| || || |   | ||||      ||         ||
Db     386785 TTGCGATTGGAAACCCAACGATTCGCCATAAGCTTTCTGCTCTTTTAGACAAGTCTGGTT 386844

Qy        266 TTGAACCAATAACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTA 325
              | ||  || | ||       |||||    |   ||  |        ||      || | ||
Db     386845 TCGATGCATTTACGGCGGTGCATCCCTCTGCAGTGATAGGAAGCCAAGTAGTGTTTGGGT 386904

Qy        326 GTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTT 385
              ||| ||| || | | |     |      || || || |||       | ||   | | |
Db     386905 CTGGGGCTGTTGTATGTTCTGGGGCCGTGATTTCGACCAATGTAAGGCTTGGTGTGCATG 386964

Qy        386 TTCATGCAAACATATACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACAT 445
              | |||   |||  | |         | |  |||||||       | | ||||||||| |
Db     386965 TGCATATCAACCCAAATGTGACAATCGGGCATGATTCAATTCTTGAAGACTATGTTTCGA 387024

Qy        446 TTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTGAAGACAATGCATATATAGGCT 505
              |   ||| |  || | |  |   ||  | ||  | ||    ||  |  |   ||| ||
Db     387025 TAAATCCGGCTGCAACAATTTCGGGTGAAGTCCTGATCATGGAAGAAACTCTTATTGGAG 387084

Qy        506 CGGGTGCAGTAATTAAGCAGGGTGTTCCTA 535
              | |  ||   | |   ||| ||  |  |||
Db     387085 CTGCGGCTACAGTATTGCAAGGACTGACTA 387114


RESULT 7
US-09-815-242-4528
; Sequence 4528, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4528
;    LENGTH: 627
;    TYPE: DNA
;    ORGANISM: Staphylococcus aureus
US-09-815-242-4528

  Query Match             6.7%;  Score 44.6;  DB 3;  Length 627;
  Best Local Similarity   54.6%;  Pred. No. 1.1;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              | |||||||||  |||||  | | ||  | || || |||  |   |||  || ||   ||
```

```
Db        356 ATATAGGAGACAATGTTTTAATAGCAGCGGGTGCGAAAGTTTTAGGAAATATTAAAATAA 415

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |    |||| | |||| ||  |   | ||||  ||   ||    ||||| | |    ||
Db        416 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 475

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587
                || ||||      ||    ||| | || |   | || |||
Db        476 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 518


RESULT 8
US-10-282-122A-8128
; Sequence 8128, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8128
;   LENGTH: 642
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-10-282-122A-8128

  Query Match             6.7%;  Score 44.6;  DB 7;  Length 642;
  Best Local Similarity   54.6%;  Pred. No. 1.1;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              | ||||||||| |||||  | | ||  | || || ||| |   ||| || ||   ||
Db        356 ATATAGGAGACAATGTTTTAATAGCAGCCGGTGCGAAAGTTTTAGGAAATATTAAAATAA 415

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |    |||| | |||| ||  |   | ||||  ||   ||    ||||| | |    ||
Db        416 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 475

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587
                || ||||      ||    ||| | || |   | || |||
Db        476 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 518


RESULT 9
US-09-815-242-8357
; Sequence 8357, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8357
;   LENGTH: 681
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(681)
US-09-815-242-8357

  Query Match             6.7%;  Score 44.6;  DB 3;  Length 681;
  Best Local Similarity   54.6%;  Pred. No. 1.2;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              | ||||||||| |||||  | | ||  | || || ||| |   ||| || ||   ||
Db        395 ATATAGGAGACAATGTTTTAATAGCAGCGGGTGCGAAAGTTTTAGGAAATATTAAAATAA 454

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |    |||| | |||| ||  |   | ||||  ||   ||    ||||| | |    ||
Db        455 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 514

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587
                || ||||      ||    ||| | || |   | || |||
Db        515 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 557


RESULT 10
US-09-815-242-8685
; Sequence 8685, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
```

```
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8685
;   LENGTH: 681
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(681)
US-09-815-242-8685

  Query Match             6.7%;  Score 44.6;  DB 3;  Length 681;
  Best Local Similarity   54.6%;  Pred. No. 1.2;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              | ||||||||| |||||  | | ||  | || || ||| |   ||| || ||   ||
Db        395 ATATAGGAGACAATGTTTTAATAGCAGCCGGTGCGAAAGTTTTAGGAAATATTAAAATAA 454

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |    |||| | |||| ||  |   | ||||  ||   ||    |||||| | |    ||
Db        455 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 514

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587
              || ||||       ||    ||| | || |   | || |||
Db        515 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 557


RESULT 11
US-10-857-625-15
; Sequence 15, Application US/10857625
; Publication No. US20050026189A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  TITLE OF INVENTION: MICROBIAL OPERONS
;  FILE REFERENCE: ELITRA.036A
;  CURRENT APPLICATION NUMBER: US/10/857,625
;  CURRENT FILING DATE:  2004-05-28
;  PRIOR APPLICATION NUMBER: 60/474768
;  PRIOR FILING DATE: 2003-05-29
;  NUMBER OF SEQ ID NOS: 833
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
;   LENGTH: 3700
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-10-857-625-15

  Query Match             6.7%;  Score 44.6;  DB 8;  Length 3700;
  Best Local Similarity   54.6%;  Pred. No. 2.6;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              | ||||||||| |||||  | | ||  | || || ||| |   ||| || ||   ||
Db        356 ATATAGGAGACAATGTTTTAATAGCAGCCGGTGCGAAAGTTTTAGGAAATATTAAAATAA 415

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |    |||| | |||| ||  |   | ||||  ||   ||    |||||| | |    ||
Db        416 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 475

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587
              || ||||       ||    ||| | || |   | || |||
Db        476 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 518


RESULT 12
US-10-857-625-14
; Sequence 14, Application US/10857625
; Publication No. US20050026189A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  TITLE OF INVENTION: MICROBIAL OPERONS
;  FILE REFERENCE: ELITRA.036A
;  CURRENT APPLICATION NUMBER: US/10/857,625
;  CURRENT FILING DATE:  2004-05-28
;  PRIOR APPLICATION NUMBER: 60/474768
;  PRIOR FILING DATE: 2003-05-29
;  NUMBER OF SEQ ID NOS: 833
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
;   LENGTH: 4350
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-10-857-625-14

  Query Match             6.7%;  Score 44.6;  DB 8;  Length 4350;
  Best Local Similarity   54.6%;  Pred. No. 2.8;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              | ||||||||| |||||  | | ||  | || || ||| |   ||| || ||   ||
Db        356 ATATAGGAGACAATGTTTTAATAGCAGCCGGTGCGAAAGTTTTAGGAAATATTAAAATAA 415

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |    |||| | |||| ||  |   | ||||  ||   ||    |||||| | |    ||
Db        416 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 475

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587
              || ||||       ||    ||| | || |   | || |||
Db        476 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 518


RESULT 13
US-08-781-986A-186/c
; Sequence 186, Application US/08781986A
; Publication No. US20030054436A1
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Staphylococcus aureus Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  5255
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
```

```
;       OPERATING SYSTEM:  MSDOS version 6.2
;       SOFTWARE:  ASCII Text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/781,986A
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Benson, Bob
;       REGISTRATION NUMBER:  30,446
;       REFERENCE/DOCKET NUMBER:  PB248PP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (301) 309-8504
;       TELEFAX:  (301) 309-8512
;   INFORMATION FOR SEQ ID NO:  186:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6876 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-781-986A-186

  Query Match             6.7%;  Score 44.6;  DB 2;  Length 6876;
  Best Local Similarity   54.6%;  Pred. No. 3.4;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              | |||||||||| |||||  | | ||  | || || ||| |   ||| || ||   ||
Db       1598 ATATAGGAGACAATGTTTTAATAGCAGCCGGTGCGAAAGTTTTAGGAAATATTAAAATAA 1539

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |    |||| | |||| ||  |   | ||||  ||   ||     ||||| | |    ||
Db       1538 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 1479

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587
                || ||||      ||    ||| | || |   | || |||
Db       1478 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 1436


RESULT 14
US-10-329-624-186/c
; Sequence 186, Application US/10329624
; Publication No. US20040043037A1
;   GENERAL INFORMATION:
;       APPLICANT: Charles Kunsch
;                  Gil H. Choi
;                  Patrick S. Dillon
;                  Craig A. Rosen
;                  Steven C. Barash
;                  Michael R. Fannon
;       TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;       NUMBER OF SEQUENCES: 5256
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: Maryland
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;            COMPUTER: HP Vectra 486/33
;            OPERATING SYSTEM: MSDOS version 6.2
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/329,624
;            FILING DATE: 27-Dec-2002
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/956,171
;            FILING DATE: October 20, 1997
;            APPLICATION NUMBER: 60/009,861
;            FILING DATE: January 5, 1996
;            APPLICATION NUMBER: 08/781,986
;            FILING DATE: January 3, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Mark J. Hyman
;            REGISTRATION NUMBER: 46,789
;            REFERENCE/DOCKET NUMBER: PB248P1D1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (240) 314-1224
;            TELEFAX: (301) 309-8439
;   INFORMATION FOR SEQ ID NO: 186:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 6876 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 186:
US-10-329-624-186

  Query Match             6.7%;  Score 44.6;  DB 7;  Length 6876;
  Best Local Similarity   54.6%;  Pred. No. 3.4;
  Matches   89;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        425 AAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGTAATGGATATGTTGTTATTG 484
              | |||||||||| |||||  | | ||  | || || ||| |   ||| || ||   ||
Db       1598 ATATAGGAGACAATGTTTTAATAGCAGCCGGTGCGAAAGTTTTAGGAAATATTAAAATAA 1539

Qy        485 AAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAGGGTGTTCCTAATCGCCCAC 544
              |    |||| | |||| ||  |   | ||||  ||   ||     ||||| | |    ||
Db       1538 ATTCAAATGTAAATATTGGTGCAAATTCAGTTGTTTTACAATCAGTTCCAAGTTATTCAA 1479

Qy        545 TTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGT 587
                || ||||      ||    ||| | || |   | || |||
Db       1478 CGGTTGTTGGTATACCAGGACATATTGTTAAGCAAGATGGTGT 1436


RESULT 15
US-10-724-972A-460/c
; Sequence 460, Application US/10724972A
; Publication No. US20040147734A1
; GENERAL INFORMATION:
;  APPLICANT: Doucette-Stamm, Lynn
;  APPLICANT:  Bush, David
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PATH03-16
;  CURRENT APPLICATION NUMBER: US/10/724,972A
;  CURRENT FILING DATE:  2003-12-01
;  PRIOR APPLICATION NUMBER: 09/450,969
;  PRIOR FILING DATE: 1999-11-29
;  PRIOR APPLICATION NUMBER: 09/134,001
;  PRIOR FILING DATE: 1998-08-13
;  PRIOR APPLICATION NUMBER: 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 460
;   LENGTH: 972
;   TYPE: DNA
;   ORGANISM: S.epidermidis
US-10-724-972A-460

  Query Match             6.5%;  Score 43;  DB 7;  Length 972;
  Best Local Similarity   50.2%;  Pred. No. 3.5;
  Matches  106;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy         15 TATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAATAAA 74
              |||| |   || |   |   ||| | |||  | |||    ||   |  | ||  ||||||
Db        870 TATTGTACCTGATATCAAAGATGCTGGAGCGATAATCATGATAGGATGTGAAGTAATAAA 811

Qy         75 GCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAAGAA 134
                 |     || || | ||  ||| |     | ||  |||| || ||  |||   ||
```

```
Db        810 TTCACGCCCTGCTTGTAACATTGCCCCCCACTCTGGAGATGGTGGTTGTGCACCTAAACC 751

Qy        135 AGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAA 194
              |||| |  |||| |  | |  |  |     |||| |||| |  ||    ||| ||
Db        750 TAAAAATGATAATGAACTAATGTATAAAATGATTTTTCCAAAATCAACTATCATAATAAC 691

Qy        195 GTATTTTAATGTTGCTATTGCTAATGATAAG 225
              || |  ||| |   | |  |  | |||||
Db        690 AAATATAGATGGTAAAACTTTTGGTAATAAG 660
```

Search completed: May 11, 2006, 07:08:44
Job time : 448.58 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:23:58 ; Search time 344.156 Seconds
(without alignments)
7892.621 Million cell updates/sec

Title: US-09-423-093B-2_COPY_13156_13821
Perfect score: 666
Sequence: 1 atgaatttgtatggtatttt.....gatcgccaacatctatttaa 666

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	56	8.4	651	17	US-11-098-686-8784	Sequence 8784, Ap
c 2	56	8.4	39794	17	US-11-098-686-8737	Sequence 8737, Ap
3	46.2	6.9	624	10	US-10-520-820-25	Sequence 25, Appl
c 4	44.6	6.7	200	17	US-11-098-686-360	Sequence 360, App
c 5	43	6.5	720	10	US-10-793-626-1695	Sequence 1695, Ap
6	43	6.5	3572	10	US-10-793-626-3988	Sequence 3988, Ap
7	43	6.5	3806	10	US-10-793-626-4166	Sequence 4166, Ap
8	41.4	6.2	2222	11	US-10-301-480-97172	Sequence 97172, A
9	41.4	6.2	2222	12	US-10-301-480-710581	Sequence 710581,
10	40.8	6.1	2944528	18	US-11-045-004-1	Sequence 1, Appli
11	40.2	6.0	1310	7	US-09-925-065A-705471	Sequence 705471,
12	39.8	6.0	471	7	US-09-925-065A-584936	Sequence 584936,
13	39.8	6.0	471	7	US-09-925-065A-584937	Sequence 584937,
14	39.8	6.0	471	7	US-09-925-065A-584938	Sequence 584938,
15	39.8	6.0	582	12	US-10-301-480-342043	Sequence 342043,
16	39.8	6.0	582	12	US-10-301-480-955452	Sequence 955452,
17	39.8	6.0	606	7	US-09-925-065A-264058	Sequence 264058,
18	39.6	5.9	822	17	US-11-098-686-9919	Sequence 9919, Ap
c 19	39.6	5.9	1457619	17	US-11-098-686-8739	Sequence 8739, Ap
20	39.4	5.9	561	12	US-10-301-480-359385	Sequence 359385,
21	39.4	5.9	561	12	US-10-301-480-972794	Sequence 972794,
22	39.4	5.9	574	7	US-09-925-065A-282448	Sequence 282448,
c 23	39.4	5.9	602	7	US-09-925-065A-585562	Sequence 585562,
24	39	5.9	561	12	US-10-301-480-359384	Sequence 359384,
25	39	5.9	561	12	US-10-301-480-972793	Sequence 972793,
26	39	5.9	574	7	US-09-925-065A-282447	Sequence 282447,
c 27	38.6	5.8	485	7	US-09-925-065A-291861	Sequence 291861,
c 28	38.6	5.8	488	12	US-10-301-480-368739	Sequence 368739,
c 29	38.6	5.8	488	12	US-10-301-480-982148	Sequence 982148,
30	38.6	5.8	606	7	US-09-925-065A-264059	Sequence 264059,
31	38.2	5.7	582	12	US-10-301-480-342042	Sequence 342042,
32	38.2	5.7	582	12	US-10-301-480-955451	Sequence 955451,
33	38.2	5.7	606	7	US-09-925-065A-264057	Sequence 264057,
34	38.2	5.7	657	7	US-09-925-065A-781569	Sequence 781569,
35	38.2	5.7	1190	7	US-09-925-065A-692870	Sequence 692870,
36	38.2	5.7	1190	7	US-09-925-065A-692871	Sequence 692871,
37	38.2	5.7	1304	7	US-09-925-065A-722880	Sequence 722880,
38	38.2	5.7	1304	7	US-09-925-065A-722881	Sequence 722881,
39	38.2	5.7	1304	7	US-09-925-065A-722882	Sequence 722882,
40	38.2	5.7	1304	7	US-09-925-065A-722883	Sequence 722883,
41	38.2	5.7	156544	17	US-11-121-086-81	Sequence 81, Appl
42	38	5.7	597	18	US-11-079-463-140	Sequence 140, App
43	38	5.7	622	7	US-09-925-065A-281193	Sequence 281193,
44	37.8	5.7	657	7	US-09-925-065A-842715	Sequence 842715,
45	37.8	5.7	1426	18	US-11-096-568A-31755	Sequence 31755, A

ALIGNMENTS

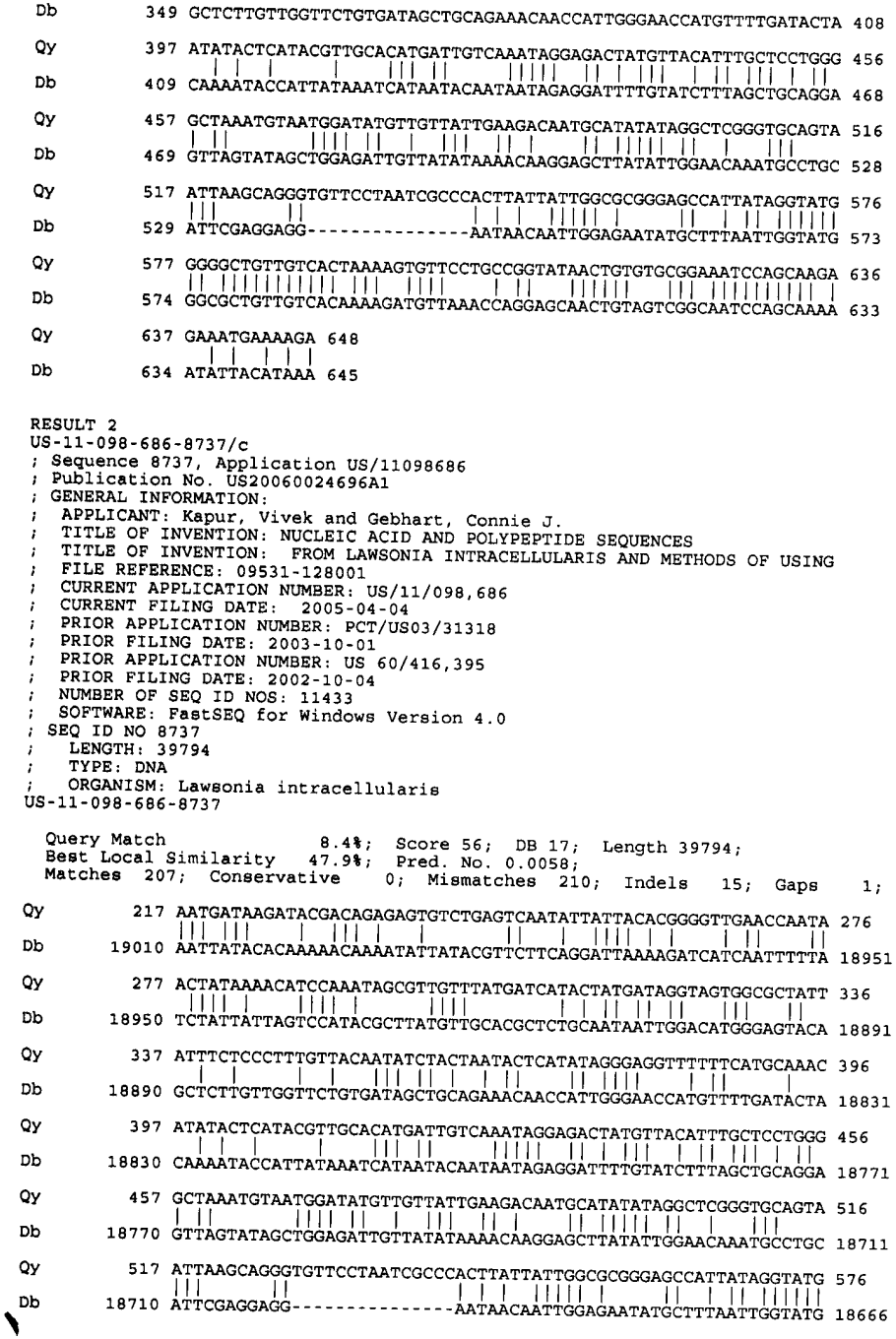
RESULT 1
US-11-098-686-8784
; Sequence 8784, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8784
; LENGTH: 651
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-8784

Query Match 8.4%; Score 56; DB 17; Length 651;
Best Local Similarity 47.9%; Pred. No. 0.002;
Matches 207; Conservative 0; Mismatches 210; Indels 15; Gaps 1;

```
Qy        217 AATGATAAGATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGGTTGAACCAATA 276
              ||| |||    |  ||| |   |        ||   |  |||| | |    | ||    ||
Db        229 AATTATACACAAAAACAAAATATTATACGTTCTTCAGGATTAAAAGATCATCAATTTTTA 288

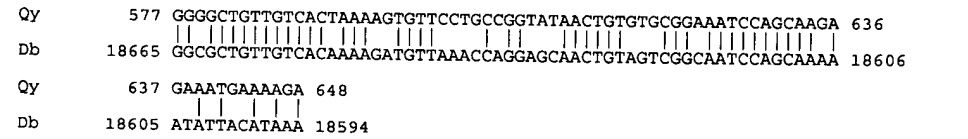
Qy        277 ACTATAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGCTATT 336
              ||||  |    ||||  |       ||||         | | || || ||   |||   ||
Db        289 TCTATTATTAGTCCATACGCTTATGTTGCACGCTCTGCAATAATTGGACATGGGAGTACA 348

Qy        337 ATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGCAAAC 396
                |  |     |  |   ||| || |  | ||    || ||||    | ||     |
```



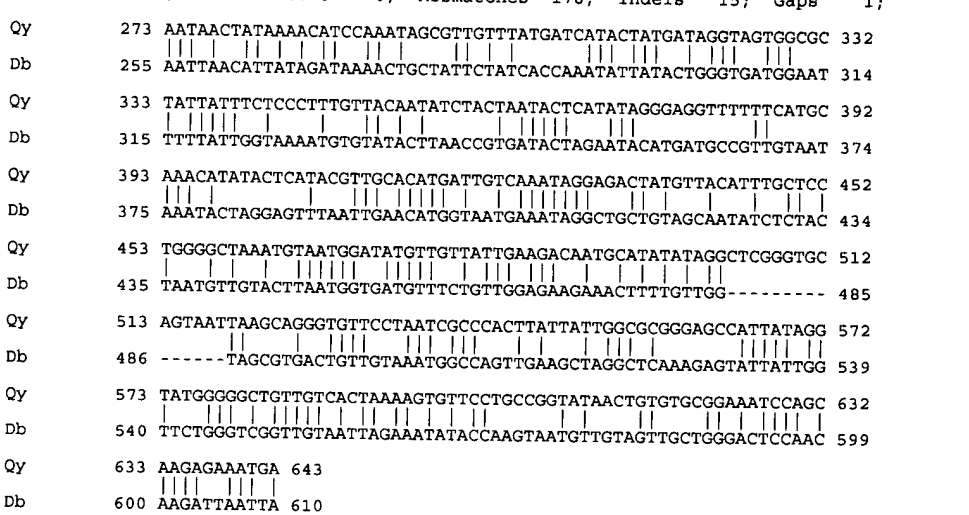
RESULT 2
US-11-098-686-8737/c
; Sequence 8737, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8737
; LENGTH: 39794
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-8737

Query Match 8.4%; Score 56; DB 17; Length 39794;
Best Local Similarity 47.9%; Pred. No. 0.0058;
Matches 207; Conservative 0; Mismatches 210; Indels 15; Gaps 1;



RESULT 3
US-10-520-820-25
; Sequence 25, Application US/10520820
; Publication No. US20060003393A1
; GENERAL INFORMATION:
; APPLICANT: MUTABILIS S.A.
; TITLE OF INVENTION: Pathogenicity determinants which can be used as targets for develo
; TITLE OF INVENTION: means for preventing and controlling bacterial infections and/or
; TITLE OF INVENTION: dissemination
; FILE REFERENCE: 1621
; CURRENT APPLICATION NUMBER: US/10/520,820
; CURRENT FILING DATE: 2005-01-07
; PRIOR APPLICATION NUMBER: PCT/EP2003/008209
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: FR 0208636
; PRIOR FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 624
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-520-820-25

Query Match 6.9%; Score 46.2; DB 10; Length 624;
Best Local Similarity 48.0%; Pred. No. 0.41;
Matches 178; Conservative 0; Mismatches 178; Indels 15; Gaps 1;



RESULT 4
US-11-098-686-360/c
; Sequence 360, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.

```
;  TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
;  TITLE OF INVENTION:  FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
;  FILE REFERENCE: 09531-128001
;  CURRENT APPLICATION NUMBER: US/11/098,686
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: PCT/US03/31318
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: US 60/416,395
;  PRIOR FILING DATE: 2002-10-04
;  NUMBER OF SEQ ID NOS: 11433
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 360
;   LENGTH: 200
;   TYPE: DNA
;   ORGANISM: Lawsonia intracellularis
US-11-098-686-360

  Query Match             6.7%;  Score 44.6;  DB 17;  Length 200;
  Best Local Similarity   65.7%;  Pred. No. 0.72;
  Matches   65;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy        543 ACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTTCC 602
              | | |  ||||| |     ||  | || |||||||| |||||||||| |||  ||||
Db         99 AATAACAATTGGAGAATATGCTTTAATTGGTATGGGCGCTGTTGTCACAAAAGATGTTAA 40

Qy        603 TGCCGGTATAACTGTGTGCGGAAATCCAGCAAGAGAAAT 641
                | ||   ||||||   ||| ||||||||||  |  | |
Db         39 ACCAGGAGCAACTGTAGTCGGCAATCCAGCAAAAATATT 1


RESULT 5
US-10-793-626-1695/c
; Sequence 1695, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1695
;   LENGTH: 720
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1695

  Query Match             6.5%;  Score 43;  DB 10;  Length 720;
  Best Local Similarity   50.2%;  Pred. No. 2.4;
  Matches  106;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy         15 TATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAATAAA 74
              |||| |   || |   |   ||| | |||  | |||    ||   | | ||  ||||||
Db        618 TATTGTACCTGATATCAAAGATGCTGGAGCGATAATCATGATAGGATGTGAAGTAATAAA 559

Qy         75 GCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAAGAA 134
                |     || || | ||  ||| |     | ||  |||| || ||  |||   ||
Db        558 TTCACGCCCTGCTTGTAACATTGCCCCCCACTCTGGAGATGGTGGTTGTGCACCTAAACC 499

Qy        135 AGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAA 194
                |||| |  |||| |  | |  |  |     |||| |||| |  ||    ||| ||
Db        498 TAAAAATGATAATGAACTAATGTATAAAATGATTTTTCCAAAATCAACTATCATAATAAC 439

Qy        195 GTATTTTAATGTTGCTATTGCTAATGATAAG 225
                || |  ||| |   | |  |  | |||||
Db        438 AAATATAGATGGTAAAACTTTTGGTAATAAG 408


RESULT 6
US-10-793-626-3988
; Sequence 3988, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3988
;   LENGTH: 3572
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3988

  Query Match             6.5%;  Score 43;  DB 10;  Length 3572;
  Best Local Similarity   50.2%;  Pred. No. 3.7;
  Matches  106;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy         15 TATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAATAAA 74
              |||| |   || |   |   ||| | |||  | |||    ||   | | ||  ||||||
Db       2178 TATTGTACCTGATATCAAAGATGCTGGAGCGATAATCATGATAGGATGTGAAGTAATAAA 2237

Qy         75 GCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAAGAA 134
                |     || || | ||  ||| |     | ||  |||| || ||  |||   ||
Db       2238 TTCACGCCCTGCTTGTAACATTGCCCCCCACTCTGGAGATGGTGGTTGTGCACCTAAACC 2297

Qy        135 AGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAA 194
                |||| |  |||| |  | |  |  |     |||| |||| |  ||    ||| ||
Db       2298 TAAAAATGATAATGAACTAATGTATAAAATGATTTTTCCAAAATCAACTATCATAATAAC 2357

Qy        195 GTATTTTAATGTTGCTATTGCTAATGATAAG 225
                || |  ||| |   | |  |  | |||||
Db       2358 AAATATAGATGGTAAAACTTTTGGTAATAAG 2388


RESULT 7
US-10-793-626-4166
; Sequence 4166, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4166
;   LENGTH: 3806
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4166

  Query Match             6.5%;  Score 43;  DB 10;  Length 3806;
  Best Local Similarity   50.2%;  Pred. No. 3.7;
  Matches  106;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy         15 TATTTTTGGTGCTGGAAGTTATGGTAGAGAAACAATACCCATTCTAAATCAACAAATAAA 74
```

```
Db        133 TATTGTACCTGATATCAAAGATGCTGGAGCGATAATCATGATAGGATGTGAAGTAATAAA 192

Qy         75 GCAAGAATGTGGTTCTGACTATGCTCTGGTTTTTGTGGATGATGTTTTGGCAGGAAAGAA 134

Db        193 TTCACGCCCTGCTTGTAACATTGCCCCCCACTCTGGAGATGGTGGTTGTGCACCTAAACC 252

Qy        135 AGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAA 194

Db        253 TAAAAATGATAATGAACTAATGTATAAAATGATTTTTCCAAAATCAACTATCATAATAAC 312

Qy        195 GTATTTTAATGTTGCTATTGCTAATGATAAG 225

Db        313 AAATATAGATGGTAAAACTTTTGGTAATAAG 343


RESULT 8
US-10-301-480-97172
; Sequence 97172, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 97172
;   LENGTH: 2222
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-97172

  Query Match             6.2%;  Score 41.4;  DB 11;  Length 2222;
  Best Local Similarity   47.4%;  Pred. No. 7.7;
  Matches  120;  Conservative    1;  Mismatches  132;  Indels    0;  Gaps    0;

Qy        221 ATAAGATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTA 280

Db        188 AAAAGCAATGTAAAAGGGAATTTGTTGCAACATAGTTTGTAGTGGCAAAAATTTAAAAAA 247

Qy        281 TAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTT 340

Db        248 AGAAAACCAACCACCAGATAGTTAATATTCTTACTATTAGTTGAAGTGGGGCTGCGGACA 307

Qy        341 CTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATAT 400

Db        308 GGACGATAGGAAAGGAGAAGAGTAGAGAATATATATGTATATTCACTACTCCATCTATAT 367

Qy        401 ACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTA 460

Db        368 ACACACACATTATATATGAACATTATATACAATACATWCATACATTGCACTATTCAGCTA 427

Qy        461 AATGTAATGGATA 473

Db        428 ACAGTAAAATATA 440


RESULT 9
US-10-301-480-710581
; Sequence 710581, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 710581
;   LENGTH: 2222
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-710581

  Query Match             6.2%;  Score 41.4;  DB 12;  Length 2222;
  Best Local Similarity   47.4%;  Pred. No. 7.7;
  Matches  120;  Conservative    1;  Mismatches  132;  Indels    0;  Gaps    0;

Qy        221 ATAAGATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTA 280

Db        188 AAAAGCAATGTAAAAGGGAATTTGTTGCAACATAGTTTGTAGTGGCAAAAATTTAAAAAA 247

Qy        281 TAAAACATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTT 340

Db        248 AGAAAACCAACCACCAGATAGTTAATATTCTTACTATTAGTTGAAGTGGGGCTGCGGACA 307

Qy        341 CTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATAT 400

Db        308 GGACGATAGGAAAGGAGAAGAGTAGAGAATATATATGTATATTCACTACTCCATCTATAT 367

Qy        401 ACTCATACGTTGCACATGATTGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTA 460

Db        368 ACACACACATTATATATGAACATTATATACAATACATWCATACATTGCACTATTCAGCTA 427

Qy        461 AATGTAATGGATA 473

Db        428 ACAGTAAAATATA 440


RESULT 10
US-11-045-004-1
; Sequence 1, Application US/11045004
; Publication No. US20060078901A1
; GENERAL INFORMATION:
;  APPLICANT: BUCHRIESER, CARMEN
;  APPLICANT:  FRANGEUL, LIONEL
;  APPLICANT:  COUVE, ELISABETH
;  APPLICANT:  RUSNIOK, CHRISTOPHE
;  APPLICANT:  FSIHI, HAFIDA
;  APPLICANT:  DEHOUX, PIERRE
;  APPLICANT:  DUSSURGET, OLIVIER
;  APPLICANT:  CHETOUANI, FARID
;  APPLICANT:  NEDJARI, HAFED
;  APPLICANT:  GLASER, PHILIPPE
;  APPLICANT:  KUNST, FRANCK
;  APPLICANT:  COSSART, PASCALE
;  APPLICANT:  DANIELS, JUSTIN
;  APPLICANT:  GOEBEI, WERNER
;  APPLICANT:  KREFT, JURGEN
;  APPLICANT:  KUHN, MICHAEL
;  APPLICANT:  NG, EVA
;  APPLICANT:  VAZQUEZ-BOLAND, ANTONIO
;  APPLICANT:  DOMINGUEZ-BERNAL, GUSTAVO
;  APPLICANT:  GARRIDO-GARCIA, PATRICIA
;  APPLICANT:  TIERREZ-MARTINEZ, ALBERTO
;  APPLICANT:  AMEND, ALEXANDRA
;  APPLICANT:  CHAKRABORTY, TRINAD
;  APPLICANT:  DOMANN, EUGEN
;  APPLICANT:  HAIN, THORSTEN
;  APPLICANT:  BERCHE, PATRICK
;  APPLICANT:  CHARBIT, ALAIN
;  APPLICANT:  DURANT, LIONEL
;  APPLICANT:  PEREZ-DIAZ, JOSE-CLAUDIO
```

```
;  APPLICANT:  BAQUERO, FERNANDO
;  APPLICANT:  GARCIA DEL PORTILLO, FRANCISCO
;  APPLICANT:  GOMEZ-LOPEZ, NURIA
;  APPLICANT:  MADUENIO, ENCARNA
;  APPLICANT:  PABLOS, BETRIZ DE
;  APPLICANT:  WEHLAND, JURGEN
;  APPLICANT:  KARST, UWE
;  APPLICANT:  ENTIAN, KARL-DIETER
;  APPLICANT:  HAUF, JORG
;  APPLICANT:  ROSE, MATTHIAS
;  APPLICANT:  VOSS, HAMUT
;  TITLE OF INVENTION: LISTERIA MONOCYTOGENES GENOME, POLYPEPTIDES AND USES
;  FILE REFERENCE: 05394.0018-02
;  CURRENT APPLICATION NUMBER: US/11/045,004
;  CURRENT FILING DATE:  2005-01-28
;  PRIOR APPLICATION NUMBER: 10/637,657
;  PRIOR FILING DATE: 2003-08-11
;  PRIOR APPLICATION NUMBER: 10/257,023
;  PRIOR FILING DATE: 2002-10-08
;  PRIOR APPLICATION NUMBER: PCT/FR01/01118
;  PRIOR FILING DATE: 2001-04-11
;  PRIOR APPLICATION NUMBER: FR 00/04,629
;  PRIOR FILING DATE: 2000-04-11
;  NUMBER OF SEQ ID NOS: 2854
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
;   LENGTH: 2944528
;   TYPE: DNA
;   ORGANISM: Listeria monocytogenes
US-11-045-004-1

  Query Match             6.1%;  Score 40.8;  DB 18;  Length 2944528;
  Best Local Similarity   63.0%;  Pred. No. 68;
  Matches   63;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        550 ATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTTGTCACTAAAAGTGTTCCTGCCGGT 609
              ||||| |  || ||  || |||    ||| ||  | || || |||  |||  || | ||
Db    1041227 ATTGGAGAAGGCGCAGTTGTAGCAGCGGGCGCAATCGTTACAAAAGATGTAGCTCCTGGA 1041286

Qy        610 ATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAAAGAT 649
              | |   |||  ||||| ||| ||| | ||| | |||| ||
Db    1041287 ACAGTCGTGGCCGGAATTCCTGCACGTGAACTTAAAAAAT 1041326


RESULT 11
US-09-925-065A-705471
; Sequence 705471, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 705471
;   LENGTH: 1310
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-705471

  Query Match             6.0%;  Score 40.2;  DB 7;  Length 1310;
  Best Local Similarity   57.6%;  Pred. No. 13;
  Matches   72;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        134 AAGTTAATGGTTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAA 193
              ||  ||||  ||  | | |     |  ||  | || |||  ||  | |||| ||  ||| |
Db        818 AAAATAATTTTTAAGGAATAGACTATACATATTGTTTTTCAATTGTCCCTCATAAAAAGA 877

Qy        194 AGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAATAT 253
              |||   | | || || | |   | | ||  ||| || ||| ||  |||    || |||||
Db        878 AGTTGCTCATTGCTGTTTTAAATTAAGAATAGAAACTACAAAGTCTGTTCCCGTAAATAT 937

Qy        254 TATTA 258
              ||||
Db        938 AATTA 942


RESULT 12
US-09-925-065A-584936
; Sequence 584936, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 584936
;   LENGTH: 471
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-584936

  Query Match             6.0%;  Score 39.8;  DB 7;  Length 471;
  Best Local Similarity   48.9%;  Pred. No. 12;
  Matches  107;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              || ||| || |||  ||| | ||   |  |||   ||      | |||     |   |||
Db         81 CCCTATATATAAATATATATAAAACCTCATATATGTAGGTTATATATATATTTGTATGTG 140

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              | || ||   |||   |||||  |  |  |  |    | ||| || ||  |  ||   |
Db        141 TTTGTGTGTGTATATATACACACGTGTATATTATATAAAATATAATATGTATGTATATTA 200

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
                |  |    ||| |||| || ||| |     |  |||     | | | |    |||
Db        201 TGTATTATATATAATATGTTATGTATTATATATTATATACACACGTGTATATATATACAC 260

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATA 399
              ||| | || || || |   | | |||     |||| | |
Db        261 ACTCAAACACACATGGAAGGCTATTTAACCACAAAAAAA 299


RESULT 13
US-09-925-065A-584937
; Sequence 584937, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
```

```
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 584937
;   LENGTH: 471
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-584937

  Query Match             6.0%;  Score 39.8;  DB 7;  Length 471;
  Best Local Similarity   48.9%;  Pred. No. 12;
  Matches  107;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              || ||| || |||  ||| | ||   |  |||   ||      | |||      |   |||
Db         81 CCCTATATATAAATATATATAAAACCTCATATATGTAGGTTATATATATATTTGTATGTG 140

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              | || ||   |||   |||||  |  |  |  |    | ||| || ||  |  ||   |
Db        141 TTTGTGTGTGTATATATACACACGTGTATATTATATAAAATATAATATGTATGTATATTA 200

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
                |  |    ||| |||| || ||| |      |  |||      | | | |    |||
Db        201 TGTATTATATATAATATGTTATGTATTATATATTATATACACACGTGTATATATATACAC 260

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATA 399
              ||| | || || || |   | | |||     |||| | |
Db        261 ACTCAAACACACATGGAAGGCTATTTAACCACAAAAAAA 299


RESULT 14
US-09-925-065A-584938
; Sequence 584938, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 584938
;   LENGTH: 471
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-584938

  Query Match             6.0%;  Score 39.8;  DB 7;  Length 471;
  Best Local Similarity   48.9%;  Pred. No. 12;
  Matches  107;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        181 CCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAGATACGACAGAGAGTG 240
              || ||| || |||  ||| | ||   |  |||   ||      | |||      |   |||
Db         81 CCCTATATATAAATATATATAAAACCTCATATATGTAGGTTATATATATATTTGTATGTG 140

Qy        241 TCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAACATCCAAATAGCGTT 300
              | || ||   |||   |||||  |  |  |  |    | ||| || ||  |  ||   |
Db        141 TTTGTGTGTGTATATATACACACGTGTATATTATATAAAATATAATATGTATGTATATTA 200

Qy        301 GTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCT 360
                |  |    ||| |||| || ||| |      |  |||      | | | |    |||
Db        201 TGTATTATATATAATATGTTATGTATTATATATTATATACACACGTGTATATATATACAC 260

Qy        361 ACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATA 399
              ||| | || || || |   | | |||     |||| | |
Db        261 ACTCAAACACACATGGAAGGCTATTTAACCACAAAAAAA 299


RESULT 15
US-10-301-480-342043
; Sequence 342043, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 342043
;   LENGTH: 582
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-342043

  Query Match             6.0%;  Score 39.8;  DB 12;  Length 582;
  Best Local Similarity   49.1%;  Pred. No. 13;
  Matches  130;  Conservative    1;  Mismatches  133;  Indels    1;  Gaps    1;

Qy        144 TTTTGAAGTGCTTTCAACCAACTGCTTTCTAAAAGCCCCTTATTTAAAAAAGTATTTTAA 203
              ||||  |||| | |  ||      |        ||  || |  | ||| ||  ||  |||| |
Db         15 TTTTACAGTGGTGTAGACATCTTCTGAGATATTAGACTTTAATTAAAGCAATGATTTGCA 74

Qy        204 TGTTGCTATTGCTAATGATAAGATACGACAGAGAGTGTCTGAGTCAATATTATTACACGG 263
              || | ||  |||| | ||  | |    ||||     | |  |  ||   |||   |||||
Db         75 TGATACTGCTGCTGAAGAATATAACATACAGCATATCTATATGTGTGTATATATACACAT 134

Qy        264 GGTTGAACCAATAACTATAAA-ACATCCAAATAGCGTTGTTTATGATCATACTATGATAG 322
              || ||  |||  ||:| | | || || |||       | |||   ||||      |||
Db        135 ATATGTACACATATATAYACATATATACACATATATACATATATACACATATATACATAT 194

Qy        323 GTAGTGGCGCTATTATTTCTCCCTTTGTTACAATATCTACTAATACTCATATAGGGAGGT 382
              ||                |  | | |      |    | || |     | || ||||||    | |
Db        195 ATACACATATATACATATATACACACATCTATACATATGTATACACACATATATACATAT 254

Qy        383 TTTTTCATGCAAACATATACTCATA 407
                |   |||  | |||||||  || |
Db        255 ATACACATATATACATATATACACA 279


Search completed: May 11, 2006, 08:30:25
Job time : 349.156 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:45:22 ; Search time 1328.5 Seconds
(without alignments)
15660.244 Million cell updates/sec

Title: US-09-423-093B-2_COPY_2744_3109
Perfect score: 366
Sequence: 1 atgataatgaataaaatcaa.....tcactttgctggctggatta 366

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	366	100.0	14002	1	AF061251	AF061251 Escherici
	2	366	100.0	14024	6	BD085555	BD085555 Nucleic a
	3	366	100.0	14029	6	AX029262	AX029262 Sequence
c	4	366	100.0	14181	6	BD184789	BD184789 Nucleic a
c	5	366	100.0	14187	6	AR204225	AR204225 Sequence
c	6	366	100.0	14187	6	AR637574	AR637574 Sequence
	7	366	100.0	31482	1	AB008676	AB008676 Escherich
c	8	366	100.0	110000	1	AE005174_28	Continuation (29 o
c	9	366	100.0	110000	1	BA000007_27	Continuation (28 o
	10	231.6	63.3	14323	1	AY730593	AY730593 Salmonell
	11	150	41.0	14224	1	AY730592	AY730592 Citrobact
	12	52.2	14.3	143794	8	AC009315	AC009315 Homo sapi
	13	50.8	13.9	109724	15	AC146910	AC146910 Medicago
c	14	50.6	13.8	19971	15	AB017059	AB017059 Arabidops
c	15	50	13.7	146597	5	CR293532	CR293532 Zebrafish
	16	50	13.7	233137	14	CR293502	CR293502 Danio rer
c	17	49.4	13.5	199814	2	CEY47D3A	AL117202 Caenorhab
	18	48	13.1	101628	8	AF130350	AF130350 Homo sapi
c	19	48	13.1	118018	8	AF227510	AF227510 Homo sapi
c	20	48	13.1	340000	8	HS21C048	AL163248 Homo sapi
	21	47.8	13.1	19822	2	AC116918	AC116918 Dictyoste
	22	47.8	13.1	250029	2	AE014820	AE014820 Plasmodiu
c	23	47.4	13.0	263335	1	CJ11168X5	AL139078 Campyloba
	24	47.2	12.9	110000	14	PFMAL7P1_11	Continuation (12 o
	25	46.8	12.8	104098	14	AP007338	AP007338 Lotus cor
c	26	46.8	12.8	222324	14	AC120334	AC120334 Rattus no
c	27	46.6	12.7	110000	15	AP008214_175	Continuation (176
c	28	46.6	12.7	144557	14	AP004631	AP004631 Oryza sat
c	29	46.6	12.7	148947	15	AP004397	AP004397 Oryza sat
	30	46.4	12.7	39984	2	AC114258	AC114258 Dictyoste
c	31	46.2	12.6	166377	14	BX957324	BX957324 Danio rer
c	32	46	12.6	109335	8	AC012449	AC012449 Homo sapi
c	33	45.8	12.5	89758	15	AC149805	AC149805 Medicago
c	34	45.8	12.5	219553	14	HSA312687	AJ312687 Homo sapi
	35	45.8	12.5	254449	2	AE014817	AE014817 Plasmodiu
	36	45.6	12.5	110000	2	AC116305_3	Continuation (4 of
c	37	45.6	12.5	110000	14	PFMAL7P1_07	Continuation (8 of
c	38	45.6	12.5	157681	14	CR339045	CR339045 Danio rer
	39	45.4	12.4	205011	5	AL954168	AL954168 Zebrafish
c	40	45.2	12.3	5542	6	AX346923	AX346923 Sequence
c	41	45.2	12.3	110000	14	PFMAL7P1_02	Continuation (3 of
	42	45.2	12.3	237549	14	AC107360	AC107360 Rattus no
	43	45.2	12.3	271929	14	AC116055	AC116055 Rattus no
c	44	45	12.3	133334	5	CR352285	CR352285 Zebrafish
c	45	45	12.3	158041	14	AC140688	AC140688 Rattus no

ALIGNMENTS

RESULT 1
AF061251
LOCUS AF061251 14002 bp DNA linear BCT 20-AUG-1998
DEFINITION Escherichia coli serotype O157:H7 O antigen gene cluster.
ACCESSION AF061251
VERSION AF061251.1 GI:3435170
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 14002)
AUTHORS Wang,L. and Reeves,P.R.
TITLE Organization of Escherichia coli O157 O antigen gene cluster and identification of its specific genes
JOURNAL Infect. Immun. 66 (8), 3545-3551 (1998)
PUBMED 9673232
REFERENCE 2 (bases 1 to 14002)
AUTHORS Wang,L.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1998) Microbiology, The University of Sydney, NSW 2006, Australia
FEATURES Location/Qualifiers
source 1. .14002
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="C664-1992"
/serotype="O157:H7"
/db_xref="taxon:562"
gene 60. .842
/gene="wbdN"
CDS 60. .842
/gene="wbdN"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdN"
/protein_id="AAC32339.1"
/db_xref="GI:3435171"
/translation="MNKETVSIIMPVYNGAKTIISSVESIIHQSYQDFVLYIIDDCST
DDTFSLINSRYKNNQKIRILRNKTNLGVAESRNYGIEMATGKYISFCDADDLWHEKKL
ERQIEVLNNECVDVVCSNYYVIDNNRNIVGEVNAPHVINYRKMLMKNYIGNLTGIYNA

```
                NKLGKFYQKKIGHEDYLMWLEIINKTNGAICIQDNLAYYMRSNNSLSGNKIKAAKWTW
                SIYREHLHLSFPKTLYYFLLYASNGVMKKITHSLLRRKETKK"
     gene       839. .2023
                /gene="wzy"
     CDS        839. .2023
                /gene="wzy"
                /codon_start=1
                /transl_table=11
                /product="O antigen polymerase Wzy"
                /protein_id="AAC32340.1"
                /db_xref="GI:3435172"
                /translation="MKSAAKLIFLFLFTLYSLQLYGVIIDDRITNFDTKVLTSIIIIF
                QIFFVLLFYLTIINERKQQKKFIVNWELKLILVFLFVTIEIAAVVLFLKEGIPIFDDD
                PGGAKLRIAEGNGLYIRYIKYFGNIVVFALIILYDEHKFKQRTIIFVYFTTIALFGYR
                SELVLLILQYILITNILSKDNRNPKIKRIIGYFLLVGVVCSLFYLSLGQDGEQNDSYN
                NMLRIINRLTIEQVEGVPYVVSESIKNDFFPTPELEKELKAIINRIQGIKHQDLFYGE
                RLHKQVFGDMGANFLSVTTYGAELLVFFGFLCVFIIPLGIYIPFYLLKRMKKTHSSIN
                CAFYSYIIMILLQYLVAGNASAFFFGPFLSVLIMCTPLILLHDTLKRLSRNENISYNC
                DL"
     gene       1992. .2738
                /gene="wbdO"
     CDS        1992. .2738
                /gene="wbdO"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase WbdO"
                /protein_id="AAC32341.1"
                /db_xref="GI:3435173"
                /translation="MKISVITVTYNNAEGLEKTLSSLSILKIKPFEIIIVDGGSTDGT
                NRVISRFTSMNITHVYEKDEGIYDAMNKGRMLAKGDLIHYLNAGDSVIGDIYKNIKEP
                CLIKVGLFENDKLLGFSSITHSNTGYCHQGVIFPKNHSEYDLRYKICADYKLIQEVFP
                EGLRSLSLITSGYVKYDMGGVSSKKRILRDKELAKIMFEKNKKNLIKFIPISIIKILF
                PERLRRVLRKMQYICLTLFFMKNSSPYDNE"
     gene       2725. .4116
                /gene="wzx"
     CDS        2725. .4116
                /gene="wzx"
                /codon_start=1
                /transl_table=11
                /product="O antigen flippase Wzx"
                /protein_id="AAC32342.1"
                /db_xref="GI:3435174"
                /translation="MIMNKIKKILKFCTLKKYDTSSALGREQERYRIISLSVISSLIS
                KILSLLSLILTVSLTLPYLGQERFGVWMTITSLGAALTFLDLGIGNALTNRIAHSFAC
                GKNLKMSRQISGGLTLLAGLSFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKV
                FVIIFGLGIYSNGVQKVYMGIQKAYISNIVNAIFILLSIITLVISSKLHAGLPVLIVS
                TLGIQYISGIYLTINLIIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSG
                DNFIISITLGVTYVAVFSITQRLFQISTVPLTIYNIPLWAAYADAHARNDTQFIKKTL
                RTSLKIVGISSFLLAFILVVFGSEVVNIWTEGKIQVPRTFIIAYALWSVIDAFSNTFA
                SFLNGLNIVKQQMLAVVTLILIAIPAKYIIVSHFGLTVMLYCFIFIYIVNYFIWYKCS
                FKKHIDRQLNIRG"
     gene       4113. .5213
                /gene="per"
     CDS        4113. .5213
                /gene="per"
                /codon_start=1
                /transl_table=11
                /product="perosamine synthetase Per"
                /protein_id="AAC32343.1"
                /db_xref="GI:3435175"
                /translation="MKMKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKF
                AEQNHVQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFV
                DSDNETWQMSVSDIEQKITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEA
                FGSKYKGKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHR
                QYWHDVIGYNYRMTNICAAIGLAQLEQADDFISRKREIADIYKKNINSLVQVHKESKD
                VFHTYWMVSILTRTAEEREELRNHLADKLIETRPVFYPVHTMPMYSEKYQKHPIAEDL
                GWRGINLPSFPSLSNEQVIYICESINEFYSDK"
     gene       5238. .6452
                /gene="wbdP"
     CDS        5238. .6452
                /gene="wbdP"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase WbdP"
                /protein_id="AAC32344.1"
                /db_xref="GI:3435176"
                /translation="MKIALNSDGFYEWGGGIDFIKYILSILETKPEICIDILLPRNDI
                HSLIREKAFPFKSILKAILKRERPRWISLNRFNEQYYRDAFTQNNIETNLTFIKSKSS
                AFYSYFDSSDCDVILPCMRVPSGNLNKKAWIGYIYDFQHCYYPSFFSKREIDQRNVFF
                KLMLNCANNIIVNAHSVITDANKYVGNYSAKLHSLPFSPCPQLKWFADYSGNIAKYNI
                DKDYFIICNQFWKHKDHATAFRAFKIYTEYNPDVYLVCTGATQDYRFPGYFNELMVLA
                KKLGIESKIKILGHIPKLEQIELIKNCIAVIQPTLFEGGPGGGVTFDAIALGKKVILS
                DIDVNKEVNCGDVYFFQAKNHYSLNDAMVKADESKIFYEPTTLIELGLKRRNACADFL
                LDVVKQEIESRS"
     gene       6472. .7590
                /gene="gmd"
     CDS        6472. .7590
                /gene="gmd"
                /codon_start=1
                /transl_table=11
                /product="GDP-D-mannose dehydratase Gmd"
                /protein_id="AAC32345.1"
                /db_xref="GI:3435177"
                /translation="MTKVALITGVTGQDGSYLAEFLLDKGYEVHGIKRRASSFNTERI
                DHIYQDPHGSNPNFHLHYGDLTDSSNLTRILKEVQPDEVYNLAAMSHVAVSFESPEYT
                ADVDAIGTLRLLEAIRFLGLENKTRFYQASTSELYGLVQEIPQKESTPFYPRSPYAVA
                KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRGLANIAQGLESCLYL
                GNMDSLRDWGHAKDYVRMQWLMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLGIKMSF
                VGKGIEEKGIVDSVEGQDAPGVKPGDVIVAVDPRYFRPAEVDTLLGDPSKANLKLGWR
                PEITLAEMISEMVAKDLEAAKKHSLLKSHGFSVSLALE"
     gene       7587. .8558
                /gene="fcl"
     CDS        7587. .8558
                /gene="fcl"
                /codon_start=1
                /transl_table=11
                /product="fucose synthetase Fcl"
                /protein_id="AAC32346.1"
                /db_xref="GI:3435178"
                /translation="MMMNKQRIFIAGHQGMVGSAITRRLKQRDDVELVLRTRDELNLL
                DSSAVLDFFSSQKIDQVYLAAAKVGGILANSSYPADFIYENIMIEANVIHAAHKNNVN
                KLLFLGSSCIYPKLAHQPIMEDELLQGKLEPTNEPYAIAKIAGIKLCESYNRQFGRDY
                RSVMPTNLYGPNDNFHPSNSHVIPALLRRFHDAVENNSPNVVVWGSGTPKREFLHVDD
                MASASIYVMEMPYDIWQKNTKVMLSHINIGTGIDCTICELAETIAKVVGYKGHITFDT
                TKPDGAPRKLLDVTLLHQLGWNHKITLHKGLENTYNWFLENQLQYRG"
     gene       8561. .9070
                /gene="wbdQ"
     CDS        8561. .9070
                /gene="wbdQ"
                /codon_start=1
                /transl_table=11
                /product="WbdQ"
                /protein_id="AAC32347.1"
                /db_xref="GI:3435179"
                /translation="MFLHSQDFATIVRSTPLISIDLIVENEFGEILLGKRINRPAQGY
                WFVPGGRVLKDEKLQTAFERLTEIELGIRLPLSVGKFYGIWQHFYEDNSMGGDFSTHY
                IVIAFLLKLQPNILKLPKSQHNAYCWLSRAKLINDDDVHYNCRAYFNNKTNDAIGLDN
                KDIICLMRQ"
     gene       9052. .10500
                /gene="manC"
     CDS        9052. .10500
                /gene="manC"
                /codon_start=1
                /transl_table=11
                /product="mannose-1-P guanosyltransferase ManC"
                /protein_id="AAC32348.1"
                /db_xref="GI:3435180"
                /translation="MSDAPIIAVVMAGGTGSRLWPLSRELYPKQFLQLSGDNTLLQTT
                LLRLSGLSCQKPLVITNEQHRFVVAEQLREINKLNGNIILEPCGRNTAPAIAISAFHA
                LKRNPQEDPLLLVLAADHVIAKESVFCDAIKNATPIANQGKIVTFGIIPEYAETGYGY
                IERGELSVPLQGHENTGFYYVNKFVEKPNRETAELYMTSGNHYWNSGIFMFKASVYLE
                ELRKFRPDIYNVCEQVASSSYIDLDFIRLSKEQFQDCPAESIDFAVMEKTEKCVVCPV
                DIGWSDVGSWQSLWDISLKSKTGDVCKGDILTYDTKNNYIYSESALVAAIGIEDMVIV
                QTKDAVLVSKKSDVQHVKKIVEMLKLQQRTEYISHREVFRPWGKFDSIDQGERYKVKK
                IIVKPGEGLSLRMHHHRSEHWIVLSGTAKVTLGDKTKLVTANESIYIPLGAAYSLENP
                GIIPLNLIEVSSGDYLGEDDIIRQKERYKHED"
```

```
     gene            10504. .11874
                     /gene="manB"
     CDS             10504. .11874
                     /gene="manB"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphomannomutase ManB"
                     /protein_id="AAC32349.1"
                     /db_xref="GI:3435181"
                     /translation="MKSLTCFKAYDIRGKLGEELNEDIAWRIGRAYGEFLKPKTIVLG
                     GDVRLTSEALKLALAKGLQDAGVDVLDIGMSGTEEIYFATFHLGVDGGIEVTASHNPM
                     DYNGMKLVREGARPISGDTGLRDVQRLAEANDFPPVDETKRGRYQQINLRDAYVDHLF
                     GYINVKNLTPLKLVINSGNGAAGPVVDAIEARFKALGAPVELIKVHNTPDGNFPNGIP
                     NPLLPECRDDTRNAVIKHGADMGIAFDGDFDRCFLFDEKGQFIEGYYIVGLLAEAFLE
                     KNPGAKIIHDPRLSWNTVDVVTAAGGTPVMSKTGHAFIKERMRKEDAIYGGEMSAHHY

  Query Match           100.0%;  Score 366;  DB 1;  Length 14002;
  Best Local Similarity  100.0%;  Pred. No. 2.9e-56;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2725 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 2784

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2785 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 2844

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2845 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 2904

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2905 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 2964

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2965 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 3024

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3025 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 3084

Qy        361 GGATTA 366
              ||||||
Db       3085 GGATTA 3090


RESULT 2
BD085555
LOCUS       BD085555               14024 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085555
VERSION     BD085555.1  GI:22631165
KEYWORDS    JP 2001522248-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 14024)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 2 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Escherichia coli
            PN   JP 2001522248-A/2
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14024
            FT                   /organism='Escherichia coli'.
FEATURES             Location/Qualifiers
     source          1. .14024
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match           100.0%;  Score 366;  DB 6;  Length 14024;
  Best Local Similarity  100.0%;  Pred. No. 2.9e-56;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2744 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 2803

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2804 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 2863

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2864 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 2923

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2924 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 2983

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2984 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 3043

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3044 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 3103

Qy        361 GGATTA 366
              ||||||
Db       3104 GGATTA 3109


RESULT 3
AX029262
LOCUS       AX029262               14029 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 2 from Patent WO9850531.
ACCESSION   AX029262
VERSION     AX029262.1  GI:10190113
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 2 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .14029
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
ORIGIN

  Query Match           100.0%;  Score 366;  DB 6;  Length 14029;
  Best Local Similarity  100.0%;  Pred. No. 2.9e-56;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
```

```
Db       2744 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 2803

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2804 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 2863

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2864 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 2923

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2924 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 2983

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2984 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 3043

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3044 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 3103

Qy        361 GGATTA 366
              ||||||
Db       3104 GGATTA 3109
```

```
RESULT 4
BD184789/c
LOCUS       BD184789               14181 bp    DNA     linear   PAT 17-JUN-2003
DEFINITION  Nucleic acid molecule and polypeptide specific to intestinal
            hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
            thereof.
ACCESSION   BD184789
VERSION     BD184789.1  GI:31876989
KEYWORDS    JP 2002355074-A/25.
SOURCE      Escherichia coli O157:H7
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 14181)
  AUTHORS   Hayashi,H., Shinagawa,H., Makino,K., Hayashi,T., Onishi,S.,
            Hattori,M. and Kurokawa,K.
  TITLE     Nucleic acid molecule and polypeptide specific to intestinal
            hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
  JOURNAL   Patent: JP 2002355074-A 25 10-DEC-2002;
            PRESIDENT OF UNIVERSITY OF TSUKUBA
COMMENT     OS   Escherichia coli O157:H7
            PN   JP 2002355074-A/25
            PD   10-DEC-2002
            PF   24-JAN-2002 JP 2002015959
            PI   HIDEO HAYASHI,HIDEO SHINAGAWA,KOZO MAKINO,TETSUYA HAYASHI,SHIN

            PI   ONISHI,
            PI   MASAHIRA HATTORI,KEN KUROKAWA
            PC   C12N15/09,C12N15/09,A61K31/7088,A61K39/00,A61K48/00,A61P31/04,
            PC   C07K14/245,
            PC   C07K16/12,C12M1/00,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/
            PC   02,C12Q1/68,
            PC   G01N33/15,G01N33/50,G01N33/53,G01N33/53,G01N33/566,G01N37/00,
            PC   C12N15/00,
            PC   C12N15/00,C12N5/00
            CC   Nucleic acid molecule and polypeptide specific to intestinal
            CC           hemorrhagic
            CC   pathogenic Escherichia coli O157:H7, and method of use thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14181
            FT                   /organism='Escherichia coli O157:H7'.
FEATURES             Location/Qualifiers
     source          1. .14181
                     /organism="Escherichia coli O157:H7"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:83334"
ORIGIN

  Query Match             100.0%;  Score 366;  DB 6;  Length 14181;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-56;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11252 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 11193

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11192 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 11133

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11132 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 11073

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11072 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 11013

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11012 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 10953

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10952 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 10893

Qy        361 GGATTA 366
              ||||||
Db      10892 GGATTA 10887
```

```
RESULT 5
AR204225/c
LOCUS       AR204225               14187 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 121 from patent US 6365723.
ACCESSION   AR204225
VERSION     AR204225.1  GI:21500820
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14187)
  AUTHORS   Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
  TITLE     Sequences of E. coli O157
  JOURNAL   Patent: US 6365723-A 121 02-APR-2002;
FEATURES             Location/Qualifiers
     source          1. .14187
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 366;  DB 6;  Length 14187;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-56;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11257 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 11198

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11197 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 11138

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11137 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 11078

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11077 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 11018

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11017 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 10958

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10957 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 10898

Qy        361 GGATTA 366
              ||||||
Db      10897 GGATTA 10892


RESULT 6
AR637574/c
LOCUS       AR637574               14187 bp    DNA     linear   PAT 20-APR-2005
DEFINITION  Sequence 121 from patent US 6855814.
ACCESSION   AR637574
VERSION     AR637574.1  GI:62771316
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14187)
  AUTHORS   Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
  TITLE     Sequences of E. coli O157
  JOURNAL   Patent: US 6855814-A 121 15-FEB-2005;
            Wisconsin Alumni Research Foundation; Madison, WI
FEATURES             Location/Qualifiers
     source          1. .14187
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match           100.0%;  Score 366;  DB 6;  Length 14187;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-56;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11257 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 11198

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11197 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 11138

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11137 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 11078

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11077 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 11018

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11017 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 10958

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10957 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 10898

Qy        361 GGATTA 366
              ||||||
Db      10897 GGATTA 10892


RESULT 7
AB008676
LOCUS       AB008676               31482 bp    DNA     linear   BCT 19-MAY-1999
DEFINITION  Escherichia coli 0157 DNA, map position at 46 min., complete cds.
ACCESSION   AB008676
VERSION     AB008676.1  GI:4867913
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Shimizu,T., Yamasaki,S., Tsukamoto,T. and Takeda,Y.
  TITLE     Analysis of the genes responsible for the O-antigen synthesis in
            enterohaemorrhagic Escherichia coli O157
  JOURNAL   Microb. Pathog. 26 (5), 235-247 (1999)
   PUBMED   10222209
REFERENCE   2  (bases 1 to 31482)
  AUTHORS   Shimizu,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1997) Takeshi Shimizu, International Medical
            Center of Japan, Department of Appropriate Technology Development,
            Research Institute; 1-21-1 Toyama, Shinjuku-ku, Tokyo 162, Japan
            (E-mail:kshimizu@imcj.go.jp, Tel:+81-3-3202-7181,
            Fax:+81-3-3202-7364)
FEATURES             Location/Qualifiers
     source          1. .31482
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="184"
                     /serotype="O157"
                     /db_xref="taxon:562"
                     /map="46 min."
                     /clone="1-4-1"
     gene            505. .1398
                     /gene="galF"
     CDS             505. .1398
                     /gene="galF"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="UTP-glucose-1-phosphate uridylyltransferase"
                     /protein_id="BAA77723.1"
                     /db_xref="GI:4867914"
                     /translation="MTNLKAVIPVAGLGMHMLPATKAIPKEMLPIVDKPMIQYIVDEI
                     VAAGIKEILLVTHASKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNV
                     RQGEPLGLGHSILCARPAIGDNPFVVVLPDVVIDDASADPLRYNLAAMIARFNETGRS
                     QVLAKRMPGDLSEYSVIQTKEPLDREGKVSRIVEFIEKPDQPQTLDSDIMAVGRYVLS
                     ADIWPELERTQPGAWGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKY
                     GLRNLKEGAKFRKGIEKLLSE"
     gene            1717. .2499
                     /gene="wbhA"
     CDS             1717. .2499
                     /gene="wbhA"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyl transferase"
                     /protein_id="BAA77724.1"
                     /db_xref="GI:4867915"
                     /translation="MNKETVSIIMPVYNGAKTIISSVESIIHQSYQDFVLYIIDDCST
                     DDTFSLINSRYKNNQKIRILRNKTNLGVAESRNYGIEMATGKYISFCDADDLWHEKKL
                     ERQIEVLNNECVDVVCSNYYVIDNKRNIVGEVNAPHVINYRKMLMKNYIGNLTGIYNA
                     NKLGKFYQKKIGHEDYLMWLEIINKTNGAICIQDNLAYYMRSNNSLSGNKIKAAKWTW
                     SIYREHLHLSFPKTLYYFLLYASNGVMKKITHSLLRRKETKK"
     gene            3153. .3680
                     /gene="wbhB"
     CDS             3153. .3680
                     /gene="wbhB"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="NADH dehydrogenase"
                     /protein_id="BAA77725.1"
                     /db_xref="GI:4867916"
                     /translation="MLRIINRLTIEQVESVPYVVSESIKNDFFPTPELEKELKAIINR
```

```
                IQGIKHQDLFYGERLHKQVFGDMGANFLSVTTYGAELLVFFGFLCVFIIPLGIYIPFY
                LLKRMKKTHSSINCAFYSYIIMILLQYLVAGNASAFFFGPFLSVLIMCTPLILLHDTL
                KRLSRNENISYNCDL"
     gene       3649. .4395
                /gene="wbhC"
     CDS        3649. .4395
                /gene="wbhC"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="glycosyl transferase"
                /protein_id="BAA77726.1"
                /db_xref="GI:4867917"
                /translation="MKISVITVTYNNAEGLEKTLSSLSILKIKPFEIIIVDGGSTDGT
                NRVISRFTSMNITHVYEKDEGIYDAMNKGRMLAKGDLIHYLNAGDSVIGDIYKNIKEP
                CLIKVGLFENDKLLGFSSITHSNTGYCHQGVIFPKNHSEYDLRYKICADYKLIQEVFP
                EGLRSLSLITSGYVKYDMGGVSSKKRILRDKELAKIMFEKNKKNLIKFIPISIIKILF
                PERLRRVLRKMQYICLTLFFMKNSSPYDNE"
     gene       4388. .5773
                /gene="wzx"
     CDS        4388. .5773
                /gene="wzx"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="flippase"
                /protein_id="BAA77727.1"
                /db_xref="GI:4867918"
                /translation="MNKIKKILKFCTLKKYDTSSALGREQERYRIISLSVISSLISKI
                LSLLSLILTVSLTLPYLGQERFGVWMTITSLGAALTFLDLGIGNALTNRIAHSFACGK
                NLKMSRQISGGLTLLAGLSFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKVFV
                IIFGLGIYSNGVQKVYMGIQKAYISNIVNAIFILLSIITLVISSKLHAGLPVLIVSTL
                GIQYISGIYLTINLIIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSGDN
                FIISITLGVTYVAVFSITQRLFQISTVPLTIYNIPLWAAYADAHARNDTQFIKKTLRT
                SLKIVGISSFLLAFILVVFGSEVVNIWTEGKIQVPRTFIIAYALWSVIDAFSNTFASF
                LNGLNIVKQQMLAVVTLILIAIPAKYIIVSHFGLTVMLYCFIFIYIVNYFIWYKCSFK
                KHIDRQLNIRG"
     gene       5776. .6870
                /gene="wbhD"
     CDS        5776. .6870
                /gene="wbhD"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="perosamine synthetase"
                /protein_id="BAA77728.1"
                /db_xref="GI:4867919"
                /translation="MKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKFAE
                QNHVQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFVDS
                DNETWQMSVSDIEQKITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEAFG
                SKYKGKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHRQY
                WHDVIGYNYRMTNICAAIGLAQLEQADDFISRKREIADIYKKNINSLVQVHKESKDVF
                HTYWMVSILTRTAEEREELRNHLADKLIETRPVFYPVHTMPMYSEKYQKHPIAEDLGW
                RGINLPSFPSLSNEQVIYICESINEFYSDK"
     gene       6895. .8109
                /gene="wbhE"
     CDS        6895. .8109
                /gene="wbhE"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="glycosyl transferase"
                /protein_id="BAA77729.1"
                /db_xref="GI:4867920"
                /translation="MKIALNSDGFYEWGGGIDFIKYILSILETKPEICIDILLPRNDI
                HSLIREKAFPPFKSILKAILKRERPRWISLNRFNEQYYRDAFTQNNIETNLTFIKSKSS
                AFYSYFDSSDCDVILPCMRVPSGNLNKKAWIGYIYDFQHCYYPSFFSKREIDQRNVFF
                KLMLNCANNIIVNAHSVITDANKYVGNYSAKLHSLPFSPCPQLKWFADYSGNIAKYNI
                DKDYFIICNQFWKHKDHATAFRAFKIYTEYNPDVYLVCTGATQDYRFPGYFNELMVLA
                KKLGIESKIKILGHIPKLEQIELIKNCIAVIQPTLFEGGPGGGVTFDAIALGKKVILS
                DIDVNKEVNCGDVYFFQAKNHYSLNDAMVKADESKIFYEPTTLIELGLKRRNACADFL
                LDVVKQEIESRS"
     gene       8129. .9247
                /gene="gmd"
     CDS        8129. .9247
                /gene="gmd"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose 4,6-dehydratase"
                /protein_id="BAA77730.1"
                /db_xref="GI:4867921"
                /translation="MTKVALITGVTGQDGSYLAEFLLDKGYEVHGIKRRASSFNTERI
                DHIYQDPHGSNPNFHLHYGDLTDSSNLTRILKEVQPDEVYNLAAMSHVAVSFESPEYT
                ADVDAIGTLRLLEAIRFLGLENKTRFYQASTSELYGLVQEIPQKESTPFYPRSPYAVA
                KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRGLANIAQGLESCLYL
                GNMDSLRDWGHAKDYVRMQWLMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLGIKMSF
                VGKGIEEKGIVDSVEGQDAPGVKPGDVIVAVDPRYFRPAEVDTLLGDPSKANLKLGWR
                PEITLAEMISEMVAKDLEAAKKHSLLKSHGFSVSLALE"
     gene       9250. .10215
                /gene="wbhF"
     CDS        9250. .10215
                /gene="wbhF"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="GDP-L-fucose pathway enzyme"
                /protein_id="BAA77731.1"
                /db_xref="GI:4867922"
                /translation="MNKQRIFIAGHQGMVGSAITRRLKQRDDVELVLRTRDELNLLDS
                SAVLDFFSSQKIDQVYLAAAKVGGILANSSYPADFIYENIMIEANVIHAAHKNNVNKL
                LFLGSSCIYPKLAHQPIMEDELLQGKLEPTNEPYAIAKIAGIKLCESYNRQFGRDYRS
                VMPTNLYGPNDNFHPSNSHVIPALLRRFHDAVENNSPNVVVWGSGTPKREFLHVDDMA
                SASIYVMEMPYDIWQKNTKVMLSHINIGTGIDCTICELAETIAKVVGYKGHITFDTTK
                PDGAPRKLLDVTLLHQLGWNHKITLHKGLENTYNWFLENQLQYRG"
     gene       10218. .10727
                /gene="wbhG"
     CDS        10218. .10727
                /gene="wbhG"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="GDP-L-fucose pathway enzyme"
                /protein_id="BAA77732.1"
                /db_xref="GI:4867923"
                /translation="MFLHSQDFATIVRSTPLISIDLIVENEFGEILLGKRINRPAQGY
                WFVPGGRVLKDEKLQTAFERLTEIELGIRLPLSVGKFYGIWQHFYEDNSMGGDFSTHY
                IVIAFLLKLQPNILKLPKSQHNAYCWLSRAKLINDDDVHYNCRAYFNNKTNDAIGLDN
                KDIICLMRQ"
     gene       10739. .12157
                /gene="manC"
     CDS        10739. .12157
                /gene="manC"
                /note="putative"
                /codon_start=1
                /transl_table=11

  Query Match          100.0%;  Score 366;  DB 1;  Length 31482;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-56;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4382 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 4441

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4442 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 4501

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4502 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 4561

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4562 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 4621
```

```
Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4622 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 4681

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4682 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 4741

Qy        361 GGATTA 366
              ||||||
Db       4742 GGATTA 4747
```

RESULT 8
AE005174_28/c
WPCOMMENT
Sequence split into 56 fragments LOCUS AE005174 Accession AE005174

Fragment Name	Begin	End
AE005174_00	1	110000
AE005174_01	100001	210000
AE005174_02	200001	310000
AE005174_03	300001	410000
AE005174_04	400001	510000
AE005174_05	500001	610000
AE005174_06	600001	710000
AE005174_07	700001	810000
AE005174_08	800001	910000
AE005174_09	900001	1010000
AE005174_10	1000001	1110000
AE005174_11	1100001	1210000
AE005174_12	1200001	1310000
AE005174_13	1300001	1410000
AE005174_14	1400001	1510000
AE005174_15	1500001	1610000
AE005174_16	1600001	1710000
AE005174_17	1700001	1810000
AE005174_18	1800001	1910000
AE005174_19	1900001	2010000
AE005174_20	2000001	2110000
AE005174_21	2100001	2210000
AE005174_22	2200001	2310000
AE005174_23	2300001	2410000
AE005174_24	2400001	2510000
AE005174_25	2500001	2610000
AE005174_26	2600001	2710000
AE005174_27	2700001	2810000
AE005174_28	2800001	2910000
AE005174_29	2900001	3010000
AE005174_30	3000001	3110000
AE005174_31	3100001	3210000
AE005174_32	3200001	3310000
AE005174_33	3300001	3410000
AE005174_34	3400001	3510000
AE005174_35	3500001	3610000
AE005174_36	3600001	3710000
AE005174_37	3700001	3810000
AE005174_38	3800001	3910000
AE005174_39	3900001	4010000
AE005174_40	4000001	4110000
AE005174_41	4100001	4210000
AE005174_42	4200001	4310000
AE005174_43	4300001	4410000
AE005174_44	4400001	4510000
AE005174_45	4500001	4610000
AE005174_46	4600001	4710000
AE005174_47	4700001	4810000
AE005174_48	4800001	4910000
AE005174_49	4900001	5010000
AE005174_50	5000001	5110000
AE005174_51	5100001	5210000
AE005174_52	5200001	5310000
AE005174_53	5300001	5410000
AE005174_54	5400001	5510000
AE005174_55	5500001	5528445

Continuation (29 of 56) of AE005174 from base 2800001 (AE005174 Escherichia coli O157:H7

Query Match 100.0%; Score 366; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred. No. 1.4e-56;
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54823 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 54764

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54763 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 54704

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54703 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 54644

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54643 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 54584

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54583 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 54524

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54523 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 54464

Qy        361 GGATTA 366
              ||||||
Db      54463 GGATTA 54458
```

RESULT 9
BA000007_27/c
WPCOMMENT
Sequence split into 55 fragments LOCUS BA000007 Accession BA000007

Fragment Name	Begin	End
BA000007_00	1	110000
BA000007_01	100001	210000
BA000007_02	200001	310000
BA000007_03	300001	410000
BA000007_04	400001	510000
BA000007_05	500001	610000
BA000007_06	600001	710000
BA000007_07	700001	810000
BA000007_08	800001	910000
BA000007_09	900001	1010000
BA000007_10	1000001	1110000
BA000007_11	1100001	1210000
BA000007_12	1200001	1310000
BA000007_13	1300001	1410000
BA000007_14	1400001	1510000
BA000007_15	1500001	1610000
BA000007_16	1600001	1710000
BA000007_17	1700001	1810000
BA000007_18	1800001	1910000
BA000007_19	1900001	2010000
BA000007_20	2000001	2110000
BA000007_21	2100001	2210000
BA000007_22	2200001	2310000
BA000007_23	2300001	2410000
BA000007_24	2400001	2510000
BA000007_25	2500001	2610000
BA000007_26	2600001	2710000
BA000007_27	2700001	2810000
BA000007_28	2800001	2910000
BA000007_29	2900001	3010000
BA000007_30	3000001	3110000

```
     BA000007_31       3100001   3210000
     BA000007_32       3200001   3310000
     BA000007_33       3300001   3410000
     BA000007_34       3400001   3510000
     BA000007_35       3500001   3610000
     BA000007_36       3600001   3710000
     BA000007_37       3700001   3810000
     BA000007_38       3800001   3910000
     BA000007_39       3900001   4010000
     BA000007_40       4000001   4110000
     BA000007_41       4100001   4210000
     BA000007_42       4200001   4310000
     BA000007_43       4300001   4410000
     BA000007_44       4400001   4510000
     BA000007_45       4500001   4610000
     BA000007_46       4600001   4710000
     BA000007_47       4700001   4810000
     BA000007_48       4800001   4910000
     BA000007_49       4900001   5010000
     BA000007_50       5000001   5110000
     BA000007_51       5100001   5210000
     BA000007_52       5200001   5310000
     BA000007_53       5300001   5410000
     BA000007_54       5400001   5498450
Continuation (28 of 55) of BA000007 from base 2700001 (BA000007 Escherichia coli O157:H7

  Query Match           100.0%;  Score 366;  DB 1;  Length 110000;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-56;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84609 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 84550

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84549 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 84490

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84489 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 84430

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84429 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 84370

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84369 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 84310

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84309 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 84250

Qy        361 GGATTA 366
              ||||||
Db      84249 GGATTA 84244


RESULT 10
AY730593
LOCUS       AY730593               14323 bp    DNA     linear   BCT 25-OCT-2004
DEFINITION  Salmonella enterica subsp. enterica serovar Urbana UDP-glucose
            pyrophosphorylase (galF) gene, partial cds; O-antigen gene cluster,
            complete sequence; and 6-phosphogluconate dehydrogenase (gnd) gene,
            partial cds.
ACCESSION   AY730593 AY012194
VERSION     AY730593.1  GI:54401420
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Urbana
  ORGANISM  Salmonella enterica subsp. enterica serovar Urbana
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 14323)
  AUTHORS   Jensen,S.O. and Reeves,P.R.
  TITLE     Molecular evolution of the GDP-mannose pathway genes (manB and
            manC) in Salmonella enterica
  JOURNAL   Microbiology (Reading, Engl.) 147 (Pt 3), 599-610 (2001)
   PUBMED   11238967
REFERENCE   2  (bases 1 to 14323)
  AUTHORS   Samuel,G., Hogbin,J.P., Wang,L. and Reeves,P.R.
  TITLE     Relationships of the Escherichia coli O157, O111, and O55 O-Antigen
            Gene Clusters with Those of Salmonella enterica and Citrobacter
            freundii, Which Express Identical O Antigens
  JOURNAL   J. Bacteriol. 186 (19), 6536-6543 (2004)
   PUBMED   15375135
REFERENCE   3  (bases 1 to 14323)
  AUTHORS   Jenkins,S.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-2000) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
REFERENCE   4  (bases 1 to 14323)
  AUTHORS   Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2004) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
  REMARK    Sequence update by submitter
COMMENT     On Oct 25, 2004 this sequence version replaced gi:11992707.
FEATURES             Location/Qualifiers
     source          1. .14323
                     /organism="Salmonella enterica subsp. enterica serovar
                     Urbana"
                     /mol_type="genomic DNA"
                     /strain="M284"
                     /serotype="30"
                     /serovar="Urbana"
                     /sub_species="enterica"
                     /db_xref="taxon:260677"
     gene            <1. .852
                     /gene="galF"
     CDS             <1. .852
                     /gene="galF"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-glucose pyrophosphorylase"
                     /protein_id="AAV34505.1"
                     /db_xref="GI:54401421"
                     /translation="MHMLPATKAIPKEMLPIVDKPMIQYIVDEIVAAGIKEIVLVTHA
                     SKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNVRQAQPLGLGHSILC
                     ARPVVGDNPFIVVLPDIIIDDATADPLRYNLAAMVARFNETGRSQVLAKRMKGDLSEY
                     SVIQTKEPLDNEGKVSRIVEFIEKPDQPQTLDSDLMAVGRYVLSADIWAELERTEPGA
                     WGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKYGLRNLKEGAKFRKS
                     IEQLLHE"
     gene            1220. .2002
                     /gene="wbdN"
     CDS             1220. .2002
                     /gene="wbdN"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAV34506.1"
                     /db_xref="GI:54401422"
                     /translation="MNREIVSVIMPVYNGSHTIVDSIESVLNQTYKDIKLYVIDDCST
                     DDTAKLLTVNYANNKNVKIIISSKNKGVAASRNLGIKESKGKYIAFCDADDLWENNKI
                     EKQINLLEHNNINVVCSNYYIMDFNKKIIGETHSPVTISYKKMLMKNYVGNLTGVYNA
                     EKLGKFYQKKVGHEDYLMWLEIIKKSNFAICLQENLAFYRVSNNSLSASKLVAAKWTW
                     NIYRECLGLSLPKSLYYFSLYAANGVIKKIGLINSDNSKAKE"
     gene            2255. .3157
                     /gene="wzy"
     CDS             2255. .3157
                     /gene="wzy"
                     /codon_start=1
                     /transl_table=11
                     /product="O-antigen polymerase"
                     /protein_id="AAV34507.1"
                     /db_xref="GI:54401423"
```

```
                /translation="MEIMAIILFAKEGIPIFDADPGGAKLKIADGNGIYIRYIKYFGN
                IVTFTLLLLLNKQKVKQKIIAIIYFLTIALFGYRSELVLLFLQYYIIMSIVNEGGNKY
                KFLNKIFIFIFMGGVIGSLFYFSLGQDGAVNNSTENIDRIINRLTIEQVESVPYIISE
                SLEYDLFPTPELGKEINAIFNRILGEKDQNLFYGERLHKKVFGDMGANFLSVTTYGAE
                LLAFFGPLCIFLIPIGLCIPFAMLKKISNTHSTVNGAIYSFLIMILLQYLIAGNASAF
                FFGPFITAIVMSIPLMILHDILKR"
gene            3163. .3894
                /gene="wbdO"
CDS             3163. .3894
                /gene="wbdO"
                /codon_start=1
                /transl_table=11
                /product="glycosyltransferase"
                /protein_id="AAV34508.1"
                /db_xref="GI:54401424"
                /translation="MNTSVITVTYNNAEGLNKTLESLCLLKVKPYEIIIIDANSIDNT
                KAIVKKYSDMNIIYICEKDNGIYDAMNKGLRLARGKLLHYLNAGDCVSGEIYADLNGP
                CLLKVGLFDKDNLIGFSSLTHSNTGYCHQGIIFPSKHPIYDLKYKLCADYKLIQMVFP
                MGLKQLPLISNGFIKYDMGGVSANKRFLRDKELLQIMLETNKKTLIKFIPLMVIKMVF
                PKKIRRILRRVNYYFSLILTNKKNV"
gene            3896. .5287
                /gene="wzx"
CDS             3896. .5287
                /gene="wzx"
                /codon_start=1
                /transl_table=11
                /product="O-antigen flippase"
                /protein_id="AAV34509.1"
                /db_xref="GI:54401425"
                /translation="MIKEKIKILFKYCTLRKYNTTTDYGRQQERYRIISLSIISNFIS
                KLLALLSLILTVSLTLPYLGQDRFGIWMTIASFSAALTFLDLGIGNALTNRIAHSFAR
                GKKIRMIRQISGGLTLLAILAFIIIAIFLTIANFVNWGSLLKGINDNTHDEVQFAIKL
                FAIIFGIGIFSNGVQRIYMGMQKAYISYISNAFFILISIISLFISSRLHAGIPMLILS
                TIGVQYLSGLSLTILLIYKKLITLNKIRINVKKETPYLISNGFFFFILQMGTLATWSG
                DNFIISTTLGVTYVSILSITQRLFQISTVPLTIYNAPLWAAYADAHARHDVDFIRKTL
                LGSLKIVGISSTTLALILVIFGQSIIAIWTGEELRVPQSLIVAYAVWSIIDAYSNTLS
                SFLNGLNIIKQQVFAVVTLILISIPSKYFIVSHFGLTAMLSCFIILYVTNYLIWYKCS
                FKKSIHLQLNIRG"
gene            5290. .6384
                /gene="per"
CDS             5290. .6384
                /gene="per"
                /codon_start=1
                /transl_table=11
                /product="perosamine synthetase"
                /protein_id="AAV34510.1"
                /db_xref="GI:54401426"
                /translation="MEYIPVYQPALSGKEKEYVNDCIDSTWISSKGSYIQRFEKKFAE
                QNNIQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPVFVDS
                DLQTWQISVDDIKNKITSRTKAIMCVHLYGHPCDMSPILEIAKVNNLFIIEDCAEAFG
                SKYKDKFVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAAHRQY
                WHDVIGYNYRMTNICAAIGLAQLEQAQLFIARKREIAEIYKNNLTPLIEVHQEGNDIF
                HTYWMVSILARNAEEREVLRSFLSEKLTETRPLFYPVHTMPMYSDKYQKHPVAENLGW
                RGINLPSFPTLSNEQVIYICNAINEFYKDK"
gene            6409. .7623
                /gene="wbdP"
CDS             6409. .7623
                /gene="wbdP"
                /codon_start=1
                /transl_table=11
                /product="glycosyltransferase"
                /protein_id="AAV34511.1"
                /db_xref="GI:54401427"
                /translation="MKITLNTDGFNEWGGGVDFIKYLLRTLEIEQDIFLDVLLPREDI
                SSLLKRNIFPVKNMIKAILSGQTPRWKKLEGFNEQYYKHAFAEITEQSKLSFITGKSS
                AYFSHYIKSDYDIILPCMRTPAVNMKNRAWLGYIYDFQHCYYPSFFSEREIAKRNLFF
                KRMLDNANNVIVNSHAVISDAKKFVGDFSANLHALPFSPCPELNWLNDNSDLIDKYAI
                DKDYFIICNQFWKHKDHSTAFKAFKKYIEINPDVYLVCTGATQDYRFPGYFDELMCLL
                KKLGISEKVKILGHIPKLEQIELIKNSIAVIQPTLFEGGPGGGITFDAVALGKKIVLS
                DIEVNKDVDCGHVYFFHSQNEYSLTEALVKATAVDHYYDPHVLLESGIKRRKKCADFI
                LNVISKEIESRI"
gene            7647. .8765
                /gene="gmd"
CDS             7647. .8765
                /gene="gmd"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose dehydratase"
                /protein_id="AAV34512.1"
                /db_xref="GI:54401428"
                /translation="MSKVALITGVTGQDGSYLAEFLLEKGYEVHGIKRRASSFNTERV
                DHIYQDPHSCNPKFHLHYGDLTDASNLTRILQEVQPDEVYNLGAMSHVAVSFESPEYT
                ADVDAMGTLRLLEAIRFLGLEKKTRFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
                KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRAIANIAQGLESCLYL
                GNMDSLRDWGHAKDYVRMQWMMLQQEQPEDFVIATGVQYSVRQFVELAAAQLGIKLRF
                EGEGINEKGIVVSVTGHDAPGVKPGDVIVAVDPRYFRPAEVETLLGDPSKAHEKLGWK
                PEITLSEMVSEMVANDLEAAKKHSLLKSHGFNVNLSLE"
gene            8768. .9733
                /gene="fcl"
CDS             8768. .9733
                /gene="fcl"
                /codon_start=1
                /transl_table=11
                /product="fucose synthetase"
                /protein_id="AAV34513.1"
                /db_xref="GI:54401429"
                /translation="MNKQRIFVAGHRGMVGSAIVRQLAQRGDVELVLRTRDELDLLDG
                RAVQAFFAGTGIDQVYLAAAKVGGIVANNTYPADFIYENMMIESNIIHAAHLHNVNKL
                LFLGSSCIYPKLARQPMAESELLQGTLEPTNEPYAIAKIAGIKLCESYNRQYGRDYRS
                VMPTNLYGPHDNFHPDNSHVIPALLRRFHEAAQSHAPEVVVWGSGTPMREFLHVDDMA
                AASIHVMELSREVWQENTAPMLSHINVGTGVDCTISELAQTIAKVVGYQGRVVFDAAK
                PDGTPRKLLDVTRLHQLGWYHEISLEAGLAGTYQWFLENQQRFRG"
gene            9736. .10236
                /gene="gmm"
CDS             9736. .10236
                /gene="gmm"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose mannosylhydrolase"
                /protein_id="AAV34514.1"
                /db_xref="GI:54401430"
                /translation="MFLRQEDFAAVVRATPLISLDFIVENGQGEILLGQRLNRPAQGY
                WFVPGGRVRKDETLEAAFARLTQAELGVRLPLAAGTFYGVWQHFYDDNFSGEDFSTHY
                IVLGFRLRVAESDLLLPDAQHGSYRWLTPEQLLASDNVHDNSRAYFQNAPYSVIGLDK
                KDVKYV"

  Query Match             63.3%;  Score 231.6;  DB 1;  Length 14323;
  Best Local Similarity   77.0%;  Pred. No. 2.8e-32;
  Matches  282;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||| |   | |||||||||| | || |||||| |||||| |||| ||||||| | ||
Db       3896 ATGATTAAAGAAAAAATCAAAATATTATTTAAATATTGCACGTTAAGAAAATATAACACT 3955

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||| ||  |  |||||| | || ||||| || |||||||| || || ||| | || ||
Db       3956 ACAACTGACTATGGTAGACAGCAAGAAAGATATAGGATTATTTCATTATCTATCATATCG 4015

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              | ||| ||||||||| || |  | |||||||| ||||| |||||||| ||| | ||||||
Db       4016 AATTTTATTAGTAAACTATTAGCGCTACTTTCGCTTATTTTAACTGTTAGTCTGACTTTA 4075

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              || ||||| || |||||  | |||||  |||||||||| ||  |  ||  || |||| |||
Db       4076 CCATATTTGGGGCAAGATCGTTTTGGAATATGGATGACAATCGCAAGCTTTAGTGCAGCT 4135

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              | || ||||| |||||||||||| |||||||||||||||||||||||| ||||||||||||
Db       4136 TTAACTTTTTTAGACTTAGGTATTGGAAATGCATTAACAAACAGGATTGCACATTCATTT 4195

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||| ||||||||||  | |  |||| | | || |||||||| ||| | ||| | || |||
Db       4196 GCGCGTGGCAAAAAAATTAGAATGATTAGACAGATTAGTGGAGGGTTGACTCTCCTCGCT 4255

Qy        361 GGATTA 366
                 ||||
```

```
Db       4256 ATATTA 4261

RESULT 11
AY730592
LOCUS       AY730592               14224 bp    DNA     linear   BCT 02-NOV-2004
DEFINITION  Citrobacter freundii serotype 90 UDP-glucose pyrophosphorylase
            (galF) gene, partial cds; O-antigen gene cluster, complete
            sequence; and 6-phosphogluconate dehydrogenase (gnd) gene, partial
            cds.
ACCESSION   AY730592
VERSION     AY730592.2  GI:55228555
KEYWORDS    .
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1  (bases 1 to 14224)
  AUTHORS   Samuel,G., Hogbin,J.P., Wang,L. and Reeves,P.R.
  TITLE     Relationships of the Escherichia coli O157, O111, and O55 O-Antigen
            Gene Clusters with Those of Salmonella enterica and Citrobacter
            freundii, Which Express Identical O Antigens
  JOURNAL   J. Bacteriol. 186 (19), 6536-6543 (2004)
   PUBMED   15375135
REFERENCE   2  (bases 1 to 14224)
  AUTHORS   Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2004) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
REFERENCE   3  (bases 1 to 14224)
  AUTHORS   Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2004) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
  REMARK    Sequence update by submitter
COMMENT     On Nov 2, 2004 this sequence version replaced gi:54401406.
FEATURES             Location/Qualifiers
     source          1. .14224
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /strain="M1972"
                     /serotype="90"
                     /db_xref="taxon:546"
     gene            <1. .773
                     /gene="galF"
     CDS             <1. .773
                     /gene="galF"
                     /codon_start=3
                     /transl_table=11
                     /product="UDP-glucose pyrophosphorylase"
                     /protein_id="AAV34492.2"
                     /db_xref="GI:55228556"
                     /translation="FVIVAAGIKEIVLVTHASKNAVENHFDTSYELESLLEQRVKRQL
                     LAEVQSICPPGVTIMNVRQAQLLGLGHSILCARPIIGDNPFVVVLPDVVLDTASADPL
                     RYNLAAMVARFNETGRSQVLAKRMEGDLSEYSVIQTKDPLDNEGKVSRIVEFIEKPDQ
                     PQTLDSDLMAVGRYVLSADIWAELEKTAPGAWGRIQLTDAISEPAKKQSVDAILMTGE
                     SYDCGKKMGYMQAFVKYGLRNLKEGPKFRKSIEKLLAM"
     gene            1104. .1853
                     /gene="wbdN"
     CDS             1104. .1853
                     /gene="wbdN"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAV34493.1"
                     /db_xref="GI:54401408"
                     /translation="METESVSIIMPVYNGASTIVQSIESILNQTYKDIRLYIIDDCST
                     DETSILLQNLYISNDRVELIRNIKNKGVAESRNIGIEQSRGKYIAFCDADDLWGETKL
                     EKQLDILESEKVDVVCTNYYLIDLEDNIIGSVCSPGIITYKTMLGKNHIGNLTGIYNV
                     KKIGKVYQKKIGHEDYLMWLEILRKTEQAICIQENLAKYRVSKKSLSANKLIAAQWTW
                     NIYRKQLKIPIIPALYFFSQYAVQGLLKKII"
     gene            1890. .3047
                     /gene="wzy"
     CDS             1890. .3047
                     /gene="wzy"
                     /codon_start=1
                     /transl_table=11
                     /product="O-antigen polymerase"
                     /protein_id="AAV34494.1"
                     /db_xref="GI:54401409"
                     /translation="MKKSVGIISFFLFSFYVLQLYGLLIDVRIKDVNIQIIAGGLVLF
                     QILCVLFYALFIRKNLSYQSGLFFNSKKIIIIFLLTFVTIELAAVILFLKEGIPIFDA
                     DPGGAKLRIADGNGIYIRYIKYFGNIVCFSLCLLLKDRKILHYSVVIIYFSTISLFGY
                     RSELIMLMIQYYILMSIINNNKVKHSMLRKVVILFSISCIVIGLFYFSLGQEGQENNY
                     ASNVNRIFNRLTIEQVESVPYIINESMKYDLFPTSEIGKEINAIANRFQGIKDNSLFY
                     GERLHKDIFGDMGENFLSVTTFGAELLAFFGPFCFFLIPLGIYIPYFLLIKMSQTKNT
                     VLCSAYSYSVVIFLQYLIAGNFSAFFFGPILTIVIMCIPLVLLHDVLKRLN"
     gene            3047. .3781
                     /gene="wbdO"
     CDS             3047. .3781
                     /gene="wbdO"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAV34495.1"
                     /db_xref="GI:54401410"
                     /translation="MKISVVTVTYNNGNGLRKTLQSLSSIKTKPCEIIIIDAESKDNT
                     KSVVDEYFSKMNIIYINEKDNGIYDAMNKGHKIANGDYIHYLNAGDCVSGDVYANVTG
                     TCLIKVGLFNERDSFLGFSNLTHSNTGYCHQGVIFPKEHLDYDLKYKLCADYDLLQKT
                     FPEGLASLPLITDGYIKYDMSGVSANKTLLRDKELLAIMWKNNKRIIPKFMIYVLFKT
                     MIPKGCRRFLRKLKYSISVFSLNGNK"
     gene            3789. .5180
                     /gene="wzx"
     CDS             3789. .5180
                     /gene="wzx"
                     /codon_start=1
                     /transl_table=11
                     /product="O-antigen flippase"
                     /protein_id="AAV34496.1"
                     /db_xref="GI:54401411"
                     /translation="MNNLLRKIYSVLKYCTLKKFDTSTDYGRQQERYRIISLSIIANL
                     ISKFLSLLTLVLTVSLTLPYLGPARFGVWMTIASLSAALTFLDLGIGNALTNRIAHAC
                     ANGERTEVIKKTSGGIALLLLVSICITCVLYIICSVIDWRIFIKGINDNILGEVQRAI
                     KLFVVIFGISILSNGIQKIYMGMQKAYITYFFNSLLSLISIFCLIYYSSLHATIPVLI
                     LISSGIPLFTGMLLMILLIKDGFVSLNEIFHNMRNEAPHLISTGFFFFILQIGTLATW
                     SGDNFIISSTLGVAFVAILSLTQRLFQISTVPLTIYNTPLWVAYADAHARKDTHFIKK
                     TLLASFKIVGFSSLIMATFLILFGRDIISVWTGDKIDIPQSFIIAYAIWSIIDAVSNT
                     LASFLNGLGIVKQQMFAVVSLVIISIPAKYFIVSHFGLTAMLSCFIAIYIVNYIVWYT
                     ISFKKCIYSQLSI"
     gene            5188. .6285
                     /gene="per"
     CDS             5188. .6285
                     /gene="per"
                     /codon_start=1
                     /transl_table=11
                     /product="perosamine synthetase"
                     /protein_id="AAV34497.1"
                     /db_xref="GI:54401412"
                     /translation="MKYIPVYEPSLTGKEKEYVNDCLDSTWISSKGSYIQRFETEFSK
                     HISAQYATTVSNGTVAIHLALIALGISEGDEVIVPTLTYIASVNAINYTGAIPVFVDS
                     DAETWQVSIEDIRNKITNKTKAIMCVHLYGQPCDMDAILEIANEFNLFVIEDCAEAFG
                     SKYKDKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYERCVHYKGQGLAAHRQY
                     WHDVIGYNYRMTNICAAIGLAQLEQAEGFINRKRNIAEIYKKELFNSPVKVHQESDNT
                     FHTYWMVSILTRTAEEREELRNYLSENFIETRPLFYPVHTMPMYSEKYQKHPVAEDLG
                     WRGMNLPSYPGMTDEQVIYICNKIKCFYNEK"
     gene            6310. .7524
                     /gene="wbdP"
     CDS             6310. .7524
                     /gene="wbdP"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAV34498.2"
                     /db_xref="GI:55228557"
                     /translation="MKITINTDGFNEWGGGVDFIKYLLLILSTKDDVFTDIIVPKDDF
                     ASLFRRNVFPIKNAIKSLGKGKFPRWVKKDGFDEGYYHNAFSGLVDKSKLTFINSNSN
```

```
                SYFDYHAKSDNDIIFPCMRVPTNSVCDSPWIGYIYDFQHCYYPSFFSAREIKKRELYF
                REMLSSANNVIVNANSVIADAKNFIGDFSADLHALPFSPCPQQIWLDDNSNLVDKYAI
                SKDYFIVCNQFWKHKNHVVVFKAFKEYIKVNPEAYLVCTGGIQDYRFPEYFGELKALI
                QKLGISSKIKILGHIPKLEQVELIKKSIAVIQPTLFEGGPGGGATYDAVALGKKVILS
                DIDVNREVNCGDVIFFNPQDEYSLAKALFKAEEEKSVKENGDLMKLGLKRREKCADFI
                MSVIRKEIESRN"
     gene       7546. .8664
                /gene="gmd"
     CDS        7546. .8664
                /gene="gmd"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose dehydratase"
                /protein_id="AAV34499.1"
                /db_xref="GI:54401414"
                /translation="MSKVALITGVTGQDGSYLAEFLLEKGYEVHGIKRRASSFNTGRV
                DHIYQDPHLGNPKFNLHYGDLTDTSNLTRILKEVQPDEVYNLGAMSHVAVSFESPEYT
                ADVDAMGTLRLLEAIRFLGLEKKTRFYQASTSELYGLVREIPQKETTPFYPRSPYAVA
                KLYAYWITVNYRESYGMYACNGILFNHESPRRGETFVTRKITRAMANIAQGLESCLYL
                GNMDSLRDWGHAKDYVRMQWMMLQQEQPEDFVIATGVQYSVRDFVEMSAAQLGIKLSF
                EGDGVNEKGIVASVTGHDAPGVKVGDVIVEVDPRYFRPAEVETLLGDPSKAHEKLGWK
                PEITLQEMVAEMVASDLEAAKKYSLLKESGFDVNLSLE"
     gene       8667. .9632
                /gene="fcl"
     CDS        8667. .9632
                /gene="fcl"
                /codon_start=1
                /transl_table=11
                /product="fucose synthetase"
                /protein_id="AAV34500.1"
                /db_xref="GI:54401415"
                /translation="MKKQRIFVAGHRGMVGTAIVRQLSLRDNVELVLRTRDELNLLDS
                SAVQAFFATERIDQVYLAAAKVGGIVANNTYPADFIYENMMIESNIIHAAHLHNVNKL
                LFLGSSCIYPKQATQPIAESELLQGTLEPTNEPYAIAKIAGIKLCESYNRQYGRDYRS
                VMPTNLYGPHDNFHPSNSHVIPALLRRFHEAREQNAPDVVVWGSGTPMREFLHVDDMA
                AASIHVMELDSEIWQEYTQPMLSHINVGTGVDCTIRELAQTIAQVVGYKGKVVFDSSK
                PDGTPRKLLDVTRLHSLGWRHSVSLEYGLESTYQWFLENQHWCRG"
     gene       9635. .10135
                /gene="gmm"
     CDS        9635. .10135
                /gene="gmm"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose mannosylhydrolase"
                /protein_id="AAV34501.1"
                /db_xref="GI:54401416"
                /translation="MFLHHDEFATVVRSTPLVSIDLIVENADGEFLLGKRLNRPAQGY
                WFVPGGRVQKDERIENAFERLTLAELGLRLPLSVAEFYGVWQHFYDDNFSGPDFSTHY
                IVLAFHIRGQESCLCLPDEQHKIWRWFKSEELEEEKDVHVNSRAYFHIENSNVIGLDN
                KGLNHV"
     gene       10128. .11576
                /gene="manC"
     CDS        10128. .11576
                /gene="manC"
                /codon_start=1
                /transl_table=11
                /product="mannose-1-P guanosyltransferase"
                /protein_id="AAV34502.1"
                /db_xref="GI:54401417"

  Query Match             41.0%;  Score 150;  DB 1;  Length 14224;
  Best Local Similarity   64.3%;  Pred. No. 1e-17;
  Matches  225;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy          8 TGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTG 67
              |||  |||||| |   | |  | ||||  || || ||||||||||  ||||| || |  |
Db       3802 TGAGAAAAATCTATTCAGTTTTAAAATACTGTACATTAAAAAAATTCGATACTTCCACAG 3861

Qy         68 CTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGA 127
                ||  ||  |  | ||||||||||| || || |||||  | ||  |||||| | | | |
Db       3862 ATTATGGACGGCAGCAGGAAAGGTATAGAATAATATCTCTTTCAATTATTGCGAATCTAA 3921

Qy        128 TTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATT 187
```

```
                | |||||  |  | ||| |  |  |  |  |  | || || |||||||| ||||| ||||
Db       3922 TCAGTAAGTTTTTATCATTGTTGACCTTAGTGCTTACCGTTAGTTTAACCTTACCATATT 3981

Qy        188 TAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACAT 247
              | || |  |  |||||||||| |||||||| ||  | ||| |  | || || || ||||
Db       3982 TGGGGCCGGCAAGATTTGGTGTGTGGATGACGATAGCAAGTTTGAGCGCAGCGCTAACAT 4041

Qy        248 TTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTTGCGTGTG 307
              |||| ||  | || || || |||||| | |||||| | ||||||||| | |  ||   ||
Db       4042 TTTTAGATCTTGGAATTGGCAATGCACTTACAAACCGTATCGCACATGCTTGCGCTAATG 4101

Qy        308 GCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTG 357
              | | |      |  | | |    ||| |||||| ||  |  || |  ||
Db       4102 GTGAGAGAACTGAAGTAATTAAAAAAACTAGTGGCGGTATAGCTCTCTTG 4151
```

```
RESULT 12
AC009315
LOCUS       AC009315              143794 bp    DNA     linear   PRI 21-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-466K9 from 2, complete sequence.
ACCESSION   AC009315
VERSION     AC009315.5  GI:9945290
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 143794)
  AUTHORS   Threideh,J., Scherger,E. and Maupin,R.
  TITLE     The sequence of Homo sapiens BAC clone RP11-466K9
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 143794)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 143794)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 143794)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 143794)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 143794)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 30, 2000 this sequence version replaced gi:9211434.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0466K09
            --------------.

            NOTICE:
```

```
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR:  pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-290K24, 200 base pair
overlap.  Actual start of this clone is at base position 1 of
RP11-466K9; actual end is at base position 56196 of RP11-290K24.

The sequence from base position 59086 to 59456 represents the best
possible assembly of an imperfect dinucleotide repeat.  This region
contains areas of low quality and may contain ambiguous bases.  The
region from 59376 to 59397 is covered only by PCR from BAC DNA.
Size of the assembly is consistent with information from PCR as
well as digest.

There are polymorphic differences in the overlap between
H_NH0466K09 and H_NH0290K24.
FEATURES             Location/Qualifiers
     source          1. .143794
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-466K9"
                     /clone_lib="RPCI-11"
     misc_feature    1911. .2185
                     /note="CpG_island (%GC=62.2, o/e=0.83, #CpGs=24)"
     misc_feature    81788. .82098
                     /note="CpG_island (%GC=60.1, o/e=0.80, #CpGs=24)"
ORIGIN

  Query Match             14.3%;  Score 52.2;  DB 8;  Length 143794;
  Best Local Similarity   50.6%;  Pred. No. 1.3;
  Matches  126;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy         33 ATTTTGCACTTTAAAAAAATATGATACATCAAGTGCTTTAGGTAGAGAACAGGAAAGGTA 92
              |||||      || | ||||||    |||||  | || | ||  ||| ||      | || |
Db     131461 ATTTTCTTAATTTAGAAAATACAGTACATAGAATGTTATACTTAGTGAGTCATACAGTTT 131520

Qy         93 CAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAGTAAAATACTCTCACTACTTTC 152
              ||  | |||||| | ||     || |    | |||  | | |||||| || ||   |||| |
Db     131521 CAATACTATATCATAGTACCATAGTGTGACTTTTTTGAATAAAACACACTATTTACTGTA 131580

Qy        153 TCTTATATTAACTGTAAGTTTAACTTTACCTTATTTAGGACAAGAGAGATTTGGTGTATG 212
              |||| ||  |  ||||   |  | | ||  | |    |     | |      |  |   ||
Db     131581 TCTTTTAAAATATGTAGTATAGAATATATTTAAACAATTTTTAAATCTGAGTTTTTACTG 131640

Qy        213 GATGACTATTACCAGTCTTGGTGCTGCTCTGACATTTTTGGACTTAGGTATAGGAAATGC 272
                |||||    | |   | ||  | | ||| ||     |   ||  | |  ||   |
Db     131641 CTAACATATTAAGCATATATATACTTTTTTTACAATTACTAATACAGAAAAATAAATGTC 131700

Qy        273 ATTAACAAA 281
```

```
              ||| | |||
Db     131701 ATTTAAAAA 131709


RESULT 13
AC146910
LOCUS       AC146910              109724 bp    DNA     linear   PLN 20-OCT-2004
DEFINITION  Medicago truncatula clone mth2-59i21, complete sequence.
ACCESSION   AC146910
VERSION     AC146910.16  GI:54306104
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 109724)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-59i21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 109724)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 109724)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 109724)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Oct 20, 2004 this sequence version replaced gi:52627254.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .109724
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-59i21"
                     /clone_lib="Medicago truncatula BAC library H2"
ORIGIN

  Query Match             13.9%;  Score 50.8;  DB 15;  Length 109724;
  Best Local Similarity   57.2%;  Pred. No. 2.5;
  Matches  111;  Conservative    0;  Mismatches   82;  Indels    1;  Gaps    1;

Qy          6 AATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAG 65
              || |||    ||  |||  || |   |||| ||||| |||||||||| ||   ||    |
Db      53954 AAAGAAGTTAAGAAAATTAAAATGGAAATATTGCATTTTAAAAAAGTAGTCTATTCCTTT 54013

Qy         66 TGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTT 125
              | ||||| ||  | || | |   | ||  || |  | ||  | |||| ||| |||| | |
Db      54014 TTCTTTAAGTTTAAAAAATGTTATTTCCAAAATAAATTTCTATTTAGTTTTTTTAAGTAT 54073

Qy        126 GATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTA 185
              | |   |   || |||| |  | ||| | |  | ||   | |   |||  | ||  ||
Db      54074 GGTCCTTGGGAT-CTCTAAAAAATTTGTTTGTTTTTTTTTTTCTTTTTTCCATTTACTAT 54132
```

```
Qy        186 TTTAGGACAAGAGA 199
              ||||| ||||| |
Db      54133 TTTAGTTCAAGATA 54146


RESULT 14
AB017059/c
LOCUS       AB017059               19971 bp    DNA     linear   PLN 14-FEB-2004
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K13P22.
ACCESSION   AB017059 BA000015
VERSION     AB017059.1  GI:3510331
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Kaneko,T., Katoh,T., Sato,S., Nakamura,Y., Asamizu,E., Kotani,H.,
            Miyajima,N. and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. IX.
            Sequence features of the regions of 1,011,550 bp covered by
            seventeen P1 and TAC clones
  JOURNAL   DNA Res. 6 (3), 183-195 (1999)
   PUBMED   10470850
REFERENCE   2  (bases 1 to 19971)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-1998) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K13P22
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is MBG8 and the 3' clone is MCO15.
FEATURES             Location/Qualifiers
     source          1. .19971
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="K13P22"
                     /clone_lib="Mitsui TAC"
                     /ecotype="Columbia"
     CDS             588. .917
                     /note="unnamed protein product; gene_id:K13P22.1
                     unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10575.1"
                     /db_xref="GI:10177376"
                     /translation="MLKILGFSGNNKVPKDTSFELPDHVTTIERNCTCFHESKTTKEG
                     INLKAFPVKSNSFSLSVIYEKLENLVIKISLNTYSLYDFKFVSNFLDRKSRQITFLLC
                     FHIYFNY"
     CDS             complement(join(3027. .3292,3379. .4531,4610. .4936))
                     /note="unnamed protein product; contains similarity to
                     MYB-related transcription factor
                     gene_id:K13P22.2"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10576.1"
                     /db_xref="GI:10177377"
                     /translation="MIMYGGGGAGKDGGSTNHLSDGGVILKKGPWTAAEDEILAAYVR
                     ENGEGNWNAVQKNTGLARCGKSCRLRWANHLRPNLKKGSFTGDEERLIIQLHAQLGNK
                     WARMAAQLPGRTDNEIKNYWNTRLKRLLRQGLPLYPPDIIPNHQLHPHPHHQQQQQHN
                     HHHHHHQQQQQHQQMYFQPQSSQRNTPSSSPLPSPTPANAKSSSSFTFHTTTANLLHP
                     LSPHTPNTPSQLSSTPPPPPLSSPLCSPRNNQYPTLPLFALPRSQINNNNNGNFTFPR
                     PPPLLQPPSSLFAKRYNNANTPLNCINRVSTAPFSPVSRDSYTSFLTLPYPSPTAQTA
                     TYHNTNNPYSSSPSFSLNPSSSSYPTSTSSPSFLHSHYTPSSTSFHTNPVYSMKQEQL
                     PSNQIPQIDGFNNVNNFTDNERQNHNLNSSGAHRRSSSCSLLEDVFEEAEALASGGRG
                     RPPKRRQLTASLPNHNNNTNNNDNFFSVSFGHYDSSDNLCSLQDLKSKEEESLQMNTM
                     QEDIAKLLDWGSDSGEISNGQSSVVTDDNLVLDVHQLASLFPADSTAVVAATNDQHNK
                     NNNNNCSWDDMQGIR"
     CDS             complement(join(6415. .7875,7965. .8110,8222. .8534))
                     /note="unnamed protein product; gb|AAF16663.1
                     gene_id:K13P22.3
                     similar to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10577.1"
                     /db_xref="GI:10177378"
                     /translation="MGKRKFEKARLKIKRAEKELKTDEKVKPDSSVKKQVRQPFSRNG
                     LEAVGSDFSSGANLASGGASQNEPVSTQIGGHEKHSYTDVLFEGNTSLVDSLEKAEDL
                     SSGKGLFGKCGRKLSVVEEDRRATYEDSDQQGDRSESIFTTFESEIKQFVAVGLHAEH
                     AYGRSLARFAATLGPVAWKIASQRIEQALPADFKFGRGWVGEYEPLPTPVLLFETCTP
                     KEPPKKASVFSKRKSNAATKTNETLFKTPLPAKEQQGSRPVRDGNHAFPFPASIGALS
                     EGSPSFVATQVGNLKSMSQHEYRNPSQLDFVKPQNRIPQQVELNLPPPAEQTNSGSGC
                     VLENQSFGKSDTVASYRSSSDMMRNMSSTDSEHYKHQMTTNGIFPGGLRNGKVSPGVN
                     NRMFDLSTDFANQMSRTATSSQQPMRQQSQSHEEQAQIMRNFNERARTQHNSTYNHPK
                     ADAPPKISSPQSARSEDSGNASVAAARAWMSIGAGGNNKQTFENASNPKSSQISAESL
                     YNPSREHFHQQAFKPRDAEETQFHPQRNGFPFQTFVHQPVHGMMNGGYQPFQNNRPIV
                     FPQMAAPTSDFSRFHVQSQWRGGITPQVQLKQRQENFNLPPDLNIGVHSPDSPAKQSS
                     GVRVDSQQPDLALQL"
     CDS             complement(join(8608. .8721,8931. .9029,9128. .9173,
                     9270. .9448))
                     /note="unnamed protein product; dbj|BAA85417.1
                     gene_id:K13P22.4
                     similar to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10578.1"
                     /db_xref="GI:10177379"
                     /translation="MLHFLCYFVFRALIIMTNLISQSALVVVLVVLGNQCDHSSETTP
                     ILDKKSLELILDKLQKKDIYGVYAEPVDPEELPDYHDMIEHPMDFSTVRKKLANGSYS
                     TLEELESDVLLICSNAMQYNSSDTVYYKQVNHVFIFSPIFFLC"
     CDS             join(16433. .16706,16993. .17138,17474. .17704,17784. .18039,
                     18206. .18429)
                     /note="gene_id:K13P22.5"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="GDSL-motif lipase/hydrolase-like protein"
                     /protein_id="BAB10579.1"
                     /db_xref="GI:10177380"
                     /translation="MPTNNTPFLTIFLLFLGLLRFDSFPGLEAATGKLASIPGLYVFG
                     DSLVDAGNNNYLPISISKANYPHNGVDFPNKKPTGRFCNGKNAADAIAEKFGLPLPPP
                     YLSLRGLLKREKRKSAAVTGVNFASGGAGIFNSSDQKLGQAIPLSKQVNNWLSIHEEV
                     MKLEPSAAQLHLSKSLFTVVIGSNDLFDYFGSFKLRRQSNPQQYTQLMADKLKEQLKR
                     IHDSGARRFLIIGVAQIGCTPGKRAKNSTLHECDEGANMWCSLYNEALVKMLQQLKQE
                     LQGSITYTYFDNYKSLHDIISNPARYGFADVTSACCGNGELNADLPCLPLAKLCSDRT
                     KHLFWDRYGHPTEAAARTIVDLMLTDDTHYSSPITLTQLVST"
     tRNA            complement(17357. .17432)
                     /product="tRNA-Phe"
                     /note="codon recognized: UUC; gene_id:K13P22.6"
                     /evidence=not_experimental
     tRNA            complement(19186. .19261)
                     /product="tRNA-Phe"
                     /note="codon recognized: UUC; gene_id:K13P22.7"
                     /evidence=not_experimental
```

ORIGIN

Query Match 13.8%; Score 50.6; DB 15; Length 19971;
Best Local Similarity 46.9%; Pred. No. 4.9;
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              || | ||  ||  |||   |||   | | || || |  ||  |  | || |||||    |
Db      14423 ATAAGAACAAAATAAACAGAAATCGTTCATACATATCCCATATCGATAAGATATGTAGTA 14364

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              |||||    |||  || | || |  |||| ||||  ||   ||   |  |  |||| ||
Db      14363 TCAAGATACTTAATTAAAAAAAATAAAAGTTACATAATCGGATAAGTCAATACTATTGCA 14304

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              | ||  |   ||  || | | |   |||||||||| | |   | ||   |  |  |
Db      14303 ATTTACAAATATATGTTAGTTTAATATTTTTCTCTTAGAATCTATATATACATTTTTGCA 14244

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              | |||  || |  |   | | ||       |  ||  |    |      ||    |
Db      14243 GTTATTTTGTGAAATAAATTCTGTAGTTGGGACTTATTTACAATGGCTGCCACTGAATTT 14184

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              || ||     ||  || |  |||     ||   ||  |  ||   ||||| |
Db      14183 TAATGTTTGTTTTTAATAATTAAAAAAAACATCTTCGAAATAAATATTTGACATTTAACA 14124

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTA 337
                | |   |||| | |  | || |||    ||| ||
Db      14123 ATCTTTCCTAAAACTCTCCACATTAGTTACTAAAATA 14087
```

RESULT 15
CR293532/c
LOCUS CR293532 146597 bp DNA linear VRT 10-JUN-2005
DEFINITION Zebrafish DNA sequence from clone DKEY-40N15 in linkage group 23, complete sequence.
ACCESSION CR293532
VERSION CR293532.12 GI:67513906
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 146597)
AUTHORS Howden,P.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT On Jun 11, 2005 this sequence version replaced gi:67513897.
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.
Any regions longer than 1kb tagged as misc-feature 'unsure' are part of a tandem repeat of more than 10kb in length where it has not been possible to anchor the base differences between repeat copies. The region has been built up based on the repeat element to match the total size of repeat indicated by restriction digest, but repeat copies may not be in the correct order and the usual finishing criteria may not apply. DKEY-40N15 is from a Zebrafish BAC library
VECTOR: pIndigoBAC-5.
FEATURES Location/Qualifiers
source 1. .146597
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/chromosome="23"
/clone="DKEY-40N15"
/clone_lib="DanioKey"
ORIGIN

Query Match 13.7%; Score 50; DB 5; Length 146597;
Best Local Similarity 50.4%; Pred. No. 3.1;
Matches 122; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```
Qy          2 TGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACAT 61
              | ||  | | |  | | | | ||| |||||||| |   | ||| || || ||
Db      12119 TCATTTTAACTCTATTTATCATAATGCTTAAATTGTTTTCCTTACAATAACATATGTTTA 12060

Qy         62 CAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAA 121
              | |||||  | |||    |      | |  | |   || |     | | | || ||| ||
Db      12059 AACGTGCTCCATGTATTTACTGCTGAGGACAGACTGTTGTGCATATTTTTTTTTTTTAAA 12000

Qy        122 GTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTAC 181
              ||| || | |||||   |      ||||  |  |  ||  |||  |||||    |
Db      11999 TTTTTATCATTAAAACTTTTAATGCACTTGTTAATTGTTTTTTTGTTTTTTTAAAGAAAA 11940

Qy        182 CTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTC 241
              |||| | ||| || || ||  | | |  ||  ||  | |   || | ||   | | |
Db      11939 TTTATGTTGGAAAAAAGTGAAGTAGAGCTTGCTTGTTTTTACACAATATTTAAGATCTTT 11880

Qy        242 TG 243
              ||
Db      11879 TG 11878
```

Search completed: May 10, 2006, 20:33:58
Job time : 1333.5 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:42:47 ; Search time 170.854 Seconds
(without alignments)
14276.972 Million cell updates/sec

Title: US-09-423-093B-2_COPY_2744_3109
Perfect score: 366
Sequence: 1 atgataatgaataaaatcaa.....tcactttgctggctggatta 366

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	366	100.0	14024	2	AAX06749	Aax06749 E. coli O
	2	366	100.0	14024	3	AAZ56386	Aaz56386 Escherich
c	3	366	100.0	14181	10	ADC01214	Adc01214 Enterohae
c	4	366	100.0	14187	9	ACD19108	Acd19108 E. coli 0
c	5	45.2	12.3	5542	6	ABL34021	Abl34021 Human imm
c	6	44.2	12.1	2000	11	ACL35363	Acl35363 Rice stre
	7	43.8	12.0	6098	4	AAS45419	Aas45419 Chemicall
	8	43.4	11.9	430	14	ADZ71008	Adz71008 Human chr
	9	43.4	11.9	728	2	ADR02010	Adr02010 A. gossyp
	10	43.4	11.9	732	2	ADR02246	Adr02246 A. gossyp
	11	43.4	11.9	749	2	ADR02343	Adr02343 A. gossyp
	12	43.4	11.9	15768	6	ABL33170	Abl33170 Human imm
	13	42.8	11.7	9965	6	ABL33527	Abl33527 Human imm
	14	42.6	11.6	2000	11	ACL35363	Acl35363 Rice stre
c	15	42.6	11.6	41624	11	ACN44166	Acn44166 Human gen
c	16	42.4	11.6	6361	6	ABL33140	Abl33140 Human imm
c	17	42.2	11.5	1011	4	AAS40845	Aas40845 cDNA enco
	18	42.2	11.5	6656	6	ABL92309	Abl92309 Chemicall
	19	42.2	11.5	6656	6	ABL49376	Abl49376 Human pol
	20	42.2	11.5	8873	6	ABK31211	Abk31211 Signal tr
	21	42.2	11.5	8873	6	ABL70174	Abl70174 Chemicall
	22	42.2	11.5	8873	6	AAS61125	Aas61125 Human gen
	23	42	11.5	2396	14	ADW10349	Adw10349 Colon pro
	24	42	11.5	2396	14	ADW10491	Adw10491 Colon pro
	25	42	11.5	40766	4	ABL28686	Abl28686 Drosophil
	26	42	11.5	42429	4	ABL17910	Abl17910 Drosophil
	27	41.8	11.4	708	2	ADR02093	Adr02093 A. gossyp
c	28	41.8	11.4	5748	6	ABL33142	Abl33142 Human imm
	29	41.6	11.4	11694	4	AAS46697	Aas46697 Tumour su
c	30	41.4	11.3	1021	4	ABA06368	Aba06368 Human cDN
c	31	41.4	11.3	1021	6	ABV83705	Abv83705 Human pol
c	32	41.4	11.3	1350	14	ADZ71037	Adz71037 Human chr
c	33	41.4	11.3	6071	6	ABL32325	Abl32325 Human imm
c	34	41.4	11.3	6071	6	AAS61076	Aas61076 Human gen
	35	41.2	11.3	1483	14	AEB70208	Aeb70208 Helicobac
	36	41	11.2	38342	4	AAS46746	Aas46746 Tumour su
	37	41	11.2	38342	6	ABK31507	Abk31507 Signal tr
c	38	40.8	11.1	535	5	ABV55427	Abv55427 Human pro
	39	40.8	11.1	6071	6	ABL32324	Abl32324 Human imm
	40	40.8	11.1	6071	6	AAS61075	Aas61075 Human gen
c	41	40.8	11.1	6222	6	ABL32693	Abl32693 Human imm
	42	40.8	11.1	14287	6	ABN80032	Abn80032 Human che
	43	40.6	11.1	6478	4	AAS45417	Aas45417 Chemicall
	44	40.6	11.1	6478	6	ABK28270	Abk28270 DNA trans
	45	40.6	11.1	6478	6	ABN80201	Abn80201 Human che

ALIGNMENTS

RESULT 1
AAX06749
ID AAX06749 standard; DNA; 14024 BP.
XX
AC AAX06749;
XX
DT 26-APR-1999 (first entry)
XX
DE E. coli O157 antigen gene cluster.
XX
KW O antigen; O157 antigen; diagnosis; wbdN gene; wzx gene; wzy gene;
KW wbdP gene; wbdO gene; wbdR gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 79. .861
FT /*tag= a
FT /product= "sugar_transferase"
FT /note= "wbdN gene (ORF1), this region is specifically
FT claimed in Claim 6; encodes AAW88312"
FT CDS 858. .2042
FT /*tag= b
FT /product= "O_antigen-polymerase"
FT /note= "wzy gene (ORF2), this region is specifically
FT claimed in Claim 6; encodes AAW88313"
FT CDS 2011. .2757
FT /*tag= c
FT /product= "sugar_transferase"
FT /note= "wbdO gene (ORF3), this region is specifically
FT claimed in Claim 6; encodes AAW88314"
FT CDS 2744. .4135
FT /*tag= d
FT /product= "O_antigen_flippase"
FT /note= "wzx gene (ORF4), this region is specifically
FT claimed in Claim 6; encodes AAW88315"
FT CDS 4132. .5232
FT /*tag= e
FT /product= "perosamine_synthetase"
FT /note= "per gene (ORF5); encodes AAW88316"
FT CDS 5257. .6471

```
FT                   /*tag= f
FT                   /product= "sugar_transferase"
FT                   /note= "wbdP gene (ORF6), this region is specifically
FT                   claimed in Claim 6; encodes AAW88317"
FT   CDS             6491. .7609
FT                   /*tag= g
FT                   /note= "ORF7; encodes AAW88318"
FT   CDS             7606. .8578
FT                   /*tag= h
FT                   /note= "ORF8; encodes AAW88319"
FT   CDS             8580. .9089
FT                   /*tag= i
FT                   /note= "ORF9; encodes AAW88320"
FT   CDS             9071. .10519
FT                   /*tag= j
FT                   /note= "manC gene (ORF10); encodes AAW88321"
FT   CDS             10523. .11893
FT                   /*tag= k
FT                   /note= "manD gene (ORF11); encodes AAW88322"
FT   repeat_region   12007. .13123
FT                   /tag= l
FT                   /note= "H-repeat, may play a role in gene cluster
FT                   assembly"
FT   CDS             13156. .13821
FT                   /*tag= m
FT                   /product= "N-acetyl_transferase"
FT                   /note= "wbdR gene (ORF12), this region is specifically
FT                   claimed in Claim 6; encodes AAW88323"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88312, AAW88313, AAW88314, AAW88315, AAW88316, AAW88317,
DR   AAW88318, AAW88319, AAW88320, AAW88321, AAW88322, AAW88323.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 6; Fig 8; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Escherichia coli O157 O antigen. The gene cluster was
CC   obtained by PCR amplification of E. coli O157:H7 strain (C664-1992) DNA
CC   using primers (see AAX06752-63) based on O antigen gene sequences. 12
CC   Open reading frames within the gene cluster were identified on the basis
CC   of homology to known sequences. The genes encode proteins (see AAW88312-
CC   23) involved the synthesis of sugars present in the polysaccharide
CC   antigen, and in the transport or processing of polysaccharide or
CC   oligosaccharide units. The use of nucleic acid molecules derived from
CC   particular assembly and transport genes, particularly wbd (transferase),
CC   wzx (flippase) and wzy (polymerase) genes, within O antigen gene clusters
CC   improves the specificity of methods for the detection and identification
CC   of O antigens, e.g. in tests of food- or faecal- derived samples, or
CC   samples from patients. The O antigen is a major virulence factor of
CC   enteropathogenic E. coli strains that cause diarrhoea and haemorrhagic
CC   colitis
XX
SQ   Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 366;  DB 2;  Length 14024;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-78;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2744 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 2803

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2804 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 2863

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2864 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 2923

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2924 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 2983

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2984 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 3043

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3044 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 3103

Qy        361 GGATTA 366
              ||||||
Db       3104 GGATTA 3109


RESULT 2
AAZ56386
ID   AAZ56386 standard; DNA; 14024 BP.
XX
AC   AAZ56386;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:56.
XX
KW   Flagellin; fliC; antigen; detection; ds.
XX
OS   Escherichia coli.
XX
PN   WO9961458-A1.
XX
PD   02-DEC-1999.
XX
PF   21-MAY-1999;   99WO-AU000385.
XX
PR   21-MAY-1998;   98AU-00003634.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 2000-072598/06.
XX
PT   Novel nucleic acid molecule useful for the detection of flagellated
PT   bacterial strains in food, feces, etc.
XX
PS   Claim 3; Page 229-233; 245pp; English.
XX
CC   AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all or
CC   part of an Escherichia coli flagellin protein except a protein expressed
CC   by E. coli H1, H7, H12 or H48 type strains. The present invention also
CC   describes a method of detecting the presence of E. coli of a particular H
CC   serotype in a sample, comprising specifically hybridising a nucleic acid,
CC   preferably at least a pair, derived from a flagellating gene, specific
CC   for a particular flagellin gene associated with the H serotype, to any
CC   E.coli in the sample which contain the gene, and detecting any hybridised
CC   molecules, identifying the presence of that serotype in the sample. (I)
```

```
CC   are useful for: (1) detecting the presence of E. coli of H serotype in a
CC   sample by hybridising at least one or a pair of (I) to any E. coli in the
CC   sample and detecting the hybridised nucleic acid molecules; and (2) for
CC   detecting the presence of both O and H-serotypes of E. coli by
CC   hybridising at least one or a pair of (I) to any E. coli present in the
CC   sample and detecting the hybridised nucleic acid molecules. (I) is
CC   particularly useful for detecting the combination of O and H antigen.
CC   Hybridised (I) when using at least one (I) is detected by southern blot
CC   analysis and, when using a pair of (I), is detected by polymerase chain
CC   reaction (PCR). AAZ56399 to AAZ56420 represent primers used in the
CC   exemplification of the present invention
XX
SQ   Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 366;  DB 3;  Length 14024;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-78;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2744 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 2803

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2804 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 2863

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2864 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 2923

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2924 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 2983

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2984 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 3043

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3044 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 3103

Qy        361 GGATTA 366
              ||||||
Db       3104 GGATTA 3109


RESULT 3
ADC01214/c
ID   ADC01214 standard; DNA; 14181 BP.
XX
AC   ADC01214;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Enterohaemorragic E. coli 0157:H7-specific nucleic acid SEQ ID NO: 1258.
XX
KW   ds; gene; enterohaemorragic; anti-bacterial.
XX
OS   Escherichia coli; 0157:H7.
XX
PN   JP2002355074-A.
XX
PD   10-DEC-2002.
XX
PF   24-JAN-2002; 2002JP-00015959.
XX
PR   24-JAN-2001; 2001JP-00112010.
XX
PA   (UYTS-) UNIV TSUKUBA.
XX
DR   WPI; 2003-451640/43.
XX
PT   Enterohemorragic Escherichia coli O157:H7-specific nucleic acid molecule
PT   and a polypeptide and its use, a polypeptide, a vector and a host cell.
XX
PS   Claim 2; SEQ ID NO 1258; 2067pp; Japanese.
XX
CC   The invention relates to a novel enterohaemorragic Escherichia coli
CC   O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
CC   has anti-bacterial activity. The polypeptide can be used in detection
CC   and/or treatment of O157:H7 infection. The nucleotide sequence of the
CC   genome of Enterohaemorragic E coli O157:H7 was determined. The present
CC   sequence represents an E. coli 0157:H7-specific nucleic acid of the
CC   invention.
XX
SQ   Sequence 14181 BP; 4450 A; 2856 C; 2281 G; 4594 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 366;  DB 10;  Length 14181;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-78;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11252 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 11193

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11192 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 11133

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11132 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 11073

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11072 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 11013

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11012 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 10953

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10952 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 10893

Qy        361 GGATTA 366
              ||||||
Db      10892 GGATTA 10887


RESULT 4
ACD19108/c
ID   ACD19108 standard; DNA; 14187 BP.
XX
AC   ACD19108;
XX
DT   27-OCT-2003  (revised)
DT   21-AUG-2003  (first entry)
XX
DE   E. coli 0157 unique DNA sequence OZID_121.
XX
KW   OZID; ds; acute haemorrhagic colitis; haemolytic uraemic syndrome;
KW   food poisoning.
XX
OS   Escherichia coli; strain 0157:H7.
XX
PN   US2003023075-A1.
XX
PD   30-JAN-2003.
XX
PF   01-APR-2002; 2002US-00114170.
XX
PR   04-DEC-1998;   98US-0110955P.
PR   03-DEC-1999;   99US-00453702.
XX
```

```
PA   (BLAT/) BLATTNER F R.
PA   (BURL/) BURLAND V D.
PA   (PERN/) PERNA N T.
PA   (PLUN/) PLUNKETT G.
PA   (WELC/) WELCH R.
XX
PI   Blattner FR,  Burland VD,  Perna NT,  Plunkett G,  Welch R;
XX
DR   WPI; 2003-479497/45.
XX
PT   New DNA sequences from Escherichia coli strain O157:H7, useful for
PT   detecting E. coli O157:H7 in a sample, or in designing diagnostic probes
PT   which can be used to distinguish strain O157:H7 from strain K12 using
PT   molecular techniques.
XX
PS   Claim 16; SEQ ID NO 121; 33pp; English.
XX
CC   The invention relates to an isolated DNA molecule comprising an E. coli
CC   strain 0157:H7 sequence selected from a clostridial cytotoxin-like gene,
CC   a urease gene cluster, a RTX toxin-like gene cluster, a locus of
CC   enterocyte effacement and 2 genes from its associated lymphocytic phage
CC   933W (a putative serine/threonine kinase and a tail fibre gene). E. coli
CC   0157:H7 can cause food poisoning, specifically acute haemorrhagic colitis
CC   (which can develop into haemolytic uraemic syndrome). Also included are
CC   an isolated DNA molecule comprising a nucleotide sequence identical to at
CC   least 25 contiguous nucleotides contained in DNA sequences selected from
CC   ACD18988-ACD19242 (being 255 E.coli 0157 DNA sequences which are not
CC   found in E.coli K12), a recombinant DNA construction comprising the DNA
CC   above and a method for detecting E. coli 0157:H7 (ATCC 43895) in a sample
CC   (or distinguishing between 0157 and K12) using a probe derived from one
CC   of the 255 sequences. The DNA sequences are useful in detecting E. coli
CC   O157:H7 in a sample, for the early diagnosis of humans and livestock
CC   infected with O157:H7, and in designing diagnostic probes which can be
CC   used to distinguish strain O157:H7 from strain K12 using molecular
CC   techniques. The present sequence is one of the 255 E. coli 0157:H7 DNA
CC   sequence (termed OZID_1-OZID255). Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from the USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030023075 (Updated on 27-OCT-2003
CC   to standardise OS field)
XX
SQ   Sequence 14187 BP; 4451 A; 2859 C; 2282 G; 4595 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 366;  DB 9;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-78;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11257 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 11198

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11197 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 11138

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11137 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 11078

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11077 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 11018

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11017 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 10958

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10957 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 10898

Qy        361 GGATTA 366
              ||||||
Db      10897 GGATTA 10892


RESULT 5
ABL34021/c
ID   ABL34021 standard; DNA; 5542 BP.
XX
AC   ABL34021;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1994.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1994; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 5542 BP; 1715 A; 55 C; 1041 G; 2731 T; 0 U; 0 Other;

  Query Match           12.3%;  Score 45.2;  DB 6;  Length 5542;
  Best Local Similarity  54.9%;  Pred. No. 0.5;
  Matches   89;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy         12 TAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCTTT 71
              |||| | ||||||| ||  |||||||  |   | || |||| || | |  |||| |   |
Db       3843 TAAATTAAAAAAAAAACAAAAATTTTACTAATCAATAAAAAATAAAATTTCAAATCTATA 3784

Qy         72 AGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAG 131
              |  ||   | |   |||   | | |  | |   || |||   || | ||  | |
Db       3783 AAATATTTACCTTAAAAAAAAAAAAAAAAAAAAAAATTATCTACCATATAAATCTAAAAATC 3724

Qy        132 TAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTT 173
                |||| ||   ||||||||   | || ||  ||  ||  ||
Db       3723 ATAAATTCTAAAACTACTTTAATTCATCTTTCCTAAAATATT 3682
```

RESULT 6
ACL35363/c
ID ACL35363 standard; cDNA; 2000 BP.
XX
AC ACL35363;
XX
DT 02-JUN-2005 (first entry)
XX
DE Rice stress-regulated promoter SEQ ID NO:13926.
XX
KW ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW agriculture.
XX
OS Oryza sativa.
XX
PN WO2003008540-A2.
XX
PD 30-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US019668.
XX
PR 22-JUN-2001; 2001US-0300112P.
PR 24-AUG-2001; 2001US-0314662P.
PR 26-SEP-2001; 2001US-0325277P.
PR 21-NOV-2001; 2001US-0332132P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
PI Moughamer T, Provart N, Ricke D, Zhu T;
XX
DR WPI; 2003-248011/24.
XX
PT New stress-responsive nucleic acid, useful for altering the
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT stress, salt stress or osmotic stress.
XX
PS Claim 48; SEQ ID NO 13926; 89pp; English.
XX
CC The invention relates to novel abiotic stress responsive polynucleotides
CC and polypeptides. Also disclosed are vectors, expression cassettes, host
CC cells, and plants containing such polynucleotides. Also disclosed are
CC methods for using the polynucleotides and polypeptides to alter the
CC responsiveness of a plant to abiotic stress. The invention is useful in
CC agriculture. The nucleic acid is useful for determining whether a test
CC plant has been exposed to an abiotic stress condition. It is also useful
CC for selecting an agent that alters abiotic stress regulated
CC polynucleotide expression in a plant cell, and to identify a homolog or
CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC molecule and the polypeptide encoded by it are useful in altering the
CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC stress, osmotic stress or any of their combinations. The present sequence
CC is used in the exemplification of the invention
XX
SQ Sequence 2000 BP; 540 A; 344 C; 416 G; 428 T; 0 U; 272 Other;

Query Match 12.1%; Score 44.2; DB 11; Length 2000;
Best Local Similarity 11.1%; Pred. No. 0.72;
Matches 28; Conservative 126; Mismatches 99; Indels 0; Gaps 0;

Qy 5 TAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAA 64
Db 335 KAWAKAWAWAYYMYSMRYRWMYYYSWAAWKGWWRWSKTTWARRWSWAWRWKTWAKWYRAA 276

Qy 65 GTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTT 124
Db 275 YSKWKKKWWYWTRAAWMATRMWTWMMTWWWARKWTRYWWAKWTYAMWWTKYAAWAAWTAT 216

Qy 125 TGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTT 184
Db 215 WTKWATRGAWWTRWSWAGWMAAAKWYWTKYWTWRRAGWMYRYRTMKMTSWCYTWWMMRWM 156

Qy 185 ATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGA 244

Db 155 WWYWTTTWYRWRKAMRKARGRAGTMWWWRKYAWSWTWTKKRKWYMWKRTMRYWKYCCKRA 96

Qy 245 CATTTTTGGACTT 257
Db 95 WMSWCYTKSWWWW 83

RESULT 7
AAS45419
ID AAS45419 standard; DNA; 6098 BP.
XX
AC AAS45419;
XX
DT 18-DEC-2001 (first entry)
XX
DE Chemically pretreated complementary DNA associated with cell cycle #62.
XX
KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW PCR primer.
XX
OS Homo sapiens.
XX
PN WO200168911-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002945.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602751/68.
XX
PT Designing primers and probes for analyzing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle.
XX
PS Claim 1; SEQ ID NO 124; 28pp; English.
XX
CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers
XX
SQ Sequence 6098 BP; 1523 A; 128 C; 1364 G; 3083 T; 0 U; 0 Other;

Query Match 12.0%; Score 43.8; DB 4; Length 6098;
Best Local Similarity 54.7%; Pred. No. 1.1;
Matches 87; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```
Qy         72 AGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAG 131
              |||||||| | | ||| ||||   | |||| |  |  |  |||| | || ||   |  |
Db       2103 AGGTAGAGGAGAAGAATGGTAGTTGTTTATTTTGTGTTAGGTTAGGTTAATTTATTTATG 2162

Qy        132 TAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTAGG 191
              || | || |    |  ||  | | |||||||     || |   ||    ||||||  |
Db       2163 TATATTATTTGATTTGGTTATTATAATATTAAGAAATAGATATTATTATTTTTATTTTAG 2222

Qy        192 ACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCT 230
              |   |||  | ||||| ||  ||    | ||   ||| |
Db       2223 AGTCGAGGAAATTGGTTTAGAGAGATTTGTTCGAAGTTT 2261
```

RESULT 8
ADZ71008
ID ADZ71008 standard; DNA; 430 BP.
XX
AC ADZ71008;
XX
DT 14-JUL-2005 (first entry)
XX
DE Human chromosome 2 contig DNA SEQ ID NO 132.
XX
KW ds; matrix attachment region; MAR; protein production.
XX
OS Homo sapiens.
XX
PN WO2005040377-A2.
XX
PD 06-MAY-2005.
XX
PF 22-OCT-2004; 2004WO-EP011974.
XX
PR 24-OCT-2003; 2003US-0513574P.
PR 06-FEB-2004; 2004EP-00002722.
XX
PA (SELE-) SELEXIS SA.
XX
PI Mermod N, Girod PA, Bucher P, Nguyen D, Calabrese D, Saugy D;
PI Puttini S;
XX
DR WPI; 2005-333507/34.
XX
PT New purified and isolated DNA sequence having protein production
PT increasing activity comprises a bent DNA element and a binding site for a
PT DNA binding protein, useful for increasing protein production activity in
PT eukaryotic host cell.
XX
PS Disclosure; SEQ ID NO 132; 282pp; English.
XX
CC The invention relates to a purified and isolated DNA sequence having
CC protein production increasing activity comprising at least one bent DNA
CC element, and at least one binding site for a DNA binding protein. The
CC purified and isolated DNA sequence comprising a first and a second
CC isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC MAR nucleotide sequence selected from a purified and isolated DNA
CC sequence above, a purified and isolated MAR DNA above, a purified and
CC isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC sequence complementary to it, its molecular chimera, or its combinations
CC and variants, is useful for increasing protein production activity in a
CC eukaryotic host cell. The present sequence represents a human chromosome
CC 2 contig DNA.
XX
SQ Sequence 430 BP; 197 A; 7 C; 7 G; 219 T; 0 U; 0 Other;

Query Match 11.9%; Score 43.4; DB 14; Length 430;
Best Local Similarity 50.6%; Pred. No. 0.84;
Matches 130; Conservative 0; Mismatches 126; Indels 1; Gaps 1;

```
Qy          8 TGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTG 67
              | |||| |   ||| | ||| ||| ||| |  | | || ||||||| ||| || |  |
Db          3 TAAATATATGTAAATATATATTTATATTATATATTATATAAAAATATAATATATAATATA 62
```

```
Qy         68 CTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGA 127
                ||  || |  | |    |    |    ||||  ||  | | | | | ||  ||  |
Db         63 TAATATATAAACTATATATTAATATAATATATATAAACTATTATATAAATACATATTAAA 122

Qy        128 TTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATT 187
              |   | | ||  | | |    | | | ||| ||  | | ||  |||||  ||    ||||
Db        123 TATATTATAT-TTTTAATATTTATATATTAAATATAATATATATTTAATATTTATATATT 181

Qy        188 TAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACAT 247
              | | | | | ||||  | | |  || |    || ||   ||  |  | | | | ||
Db        182 AAATATATAATATATTTAATATTTATATAATATATAGCATATTTTATATTTATATTATAT 241

Qy        248 TTTTGGACTTAGGTATA 264
              |     |||  | ||
Db        242 ATAACATTTTATATTTA 258
```

RESULT 9
ADR02010
ID ADR02010 standard; DNA; 728 BP.
XX
AC ADR02010;
XX
DT 23-SEP-2004 (first entry)
XX
DE A. gossypii genomic DNA PAG1462UP.
XX
KW Filamentous funghi; ds; forensic identification; gene characterisation;
KW intergenomic comparison; chromosome mapping.
XX
OS Eremothecium gossypii.
XX
PN US6239264-B1.
XX
PD 29-MAY-2001.
XX
PF 24-DEC-1997; 97US-00998416.
XX
PR 31-DEC-1996; 97CH-00000016.
XX
PA (SYNG-) SYNGENTA PARTICIPATIONS AG.
XX
PI Philippsen P, Poehlmann R, Steiner-Lange S, Mohr C, Wendland J;
PI Knechtle P, Rebischung C;
XX
DR WPI; 1998-388120/33.
XX
PT New gene for adenylate cyclase from Ashbya gossypii - useful for
PT generating recombinant microorganisms with alteration in gene of cAMP-
PT dependent signalling pathway for increasing production of fine chemicals.
XX
PS Example 3; SEQ ID NO 702; 632pp; English.
XX
CC The invention relates to isolated DNA molecules comprising isolated
CC genomic DNA sequences from the filamentous funghi Ashbya gossypii, the
CC sequences comprising ADR01309, ADR01366, ADR01367, ADR01388, ADR01428,
CC ADR01466, ADR01629, ADR01637, ADR02057, ADR02345 and ADR02369, chosen
CC from 1047 disclosed genomic sequences. Also included is a cloning vector
CC comprising a nucleotide sequence chosen from the above sequences. The
CC novel Ashbya gossypii genomic sequences are useful for forensic
CC identification, gene characterisation, for studying gene organisation by
CC intergenomic comparison (with Saccharomyces cerevisiae), identifying
CC biosynthetic genes for selectable markers, to isolate
CC promoters/terminators/centromeres, chromosome mapping, and in identifying
CC sequences unique to Ashbya gossypii for species identification. The
CC present sequence is an A. gossypii novel genomic sequence of the
CC invention.
XX
SQ Sequence 728 BP; 283 A; 83 C; 67 G; 295 T; 0 U; 0 Other;

Query Match 11.9%; Score 43.4; DB 2; Length 728;
Best Local Similarity 48.6%; Pred. No. 0.92;
Matches 119; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              || || |   | | ||  ||| || || ||||||||   |  || ||||||    ||| |
Db        415 ATAATTAATTAAATAAATAAATAATTAATTAAATTTAAAATGTTTAAAAAAAGAAATAAA 474

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              | |  || | ||  || | |     |||  |  |  |||       |  | |  || | |
Db        475 TAATATGTTATATTTAAATAGATCAAAATTTCAACAATTTCCATTTCATTTAGTACTACC 534

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              |     || |  ||      |    |    | | |||  || ||| | |  || |  ||
Db        535 ATCACCATGACCAATTGTTACATCATTTAGTTTATTAGGTTTACTATTAACTTTAGCTTT 594

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
                |||  | |  |  |  | |    | |||   | | |||| |  | || ||  |  |
Db        595 TACTATACATGGTATTATTGGTAATATTTATCCTTTATTATTATCTTTATTAGTAGTTTT 654

Qy        241 CTGAC 245
                | ||
Db        655 ATTAC 659
```

RESULT 10
ADR02246
ID ADR02246 standard; DNA; 732 BP.
XX
AC ADR02246;
XX
DT 23-SEP-2004 (first entry)
XX
DE A. gossypii genomic DNA PAG1581UP.
XX
KW Filamentous funghi; ds; forensic identification; gene characterisation;
KW intergenomic comparison; chromosome mapping.
XX
OS Eremothecium gossypii.
XX
PN US6239264-B1.
XX
PD 29-MAY-2001.
XX
PF 24-DEC-1997; 97US-00998416.
XX
PR 31-DEC-1996; 97CH-00000016.
XX
PA (SYNG-) SYNGENTA PARTICIPATIONS AG.
XX
PI Philippsen P, Poehlmann R, Steiner-Lange S, Mohr C, Wendland J;
PI Knechtle P, Rebischung C;
XX
DR WPI; 1998-388120/33.
XX
PT New gene for adenylate cyclase from Ashbya gossypii - useful for
PT generating recombinant microorganisms with alteration in gene of cAMP-
PT dependent signalling pathway for increasing production of fine chemicals.
XX
PS Example 3; SEQ ID NO 938; 632pp; English.
XX
CC The invention relates to isolated DNA molecules comprising isolated
CC genomic DNA sequences from the filamentous funghi Ashbya gossypii, the
CC sequences comprising ADR01309, ADR01366, ADR01367, ADR01388, ADR01428,
CC ADR01466, ADR01629, ADR01637, ADR02057, ADR02345 and ADR02369, chosen
CC from 1047 disclosed genomic sequences. Also included is a cloning vector
CC comprising a nucleotide sequence chosen from the above sequences. The
CC novel Ashbya gossypii genomic sequences are useful for forensic
CC identification, gene characterisation, for studying gene organisation by
CC intergenomic comparison (with Saccharomyces cerevisiae), identifying
CC biosynthetic genes for selectable markers, to isolate
CC promoters/terminators/centromeres, chromosome mapping, and in identifying
CC sequences unique to Ashbya gossypii for species identification. The
CC present sequence is an A. gossypii novel genomic sequence of the
CC invention.
XX

SQ Sequence 732 BP; 287 A; 83 C; 66 G; 296 T; 0 U; 0 Other;

  Query Match   11.9%; Score 43.4; DB 2; Length 732;
  Best Local Similarity   48.6%; Pred. No. 0.92;
  Matches 119; Conservative   0; Mismatches 126; Indels   0; Gaps   0;

```
Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              || || |   | | ||  ||| || || ||||||||   |  || ||||||    ||| |
Db        415 ATAATTAATTAAATAAATAAATAATTAATTAAATTTAAAATGTTTAAAAAAAGAAATAAA 474

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              | |  || | ||  || | |     |||  |  |  |||       |  | |  || | |
Db        475 TAATATGTTATATTTAAATAGATCAAAATTTCAACAATTTCCATTTCATTTAGTACTACC 534

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              |     || |  ||      |    |    | | |||  || ||| | |  || |  ||
Db        535 ATCACCATGACCAATTGTTACATCATTTAGTTTATTAGGTTTACTATTAACTTTAGCTTT 594

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
                |||  | |  |  |  | |    | |||   | | |||| |  | || ||  |  |
Db        595 TACTATACATGGTATTATTGGTAATATTTATCCTTTATTATTATCTTTATTAGTAGTTTT 654

Qy        241 CTGAC 245
                | ||
Db        655 ATTAC 659
```

RESULT 11
ADR02343
ID ADR02343 standard; DNA; 749 BP.
XX
AC ADR02343;
XX
DT 23-SEP-2004 (first entry)
XX
DE A. gossypii genomic DNA PAG1633RP.
XX
KW Filamentous funghi; ds; forensic identification; gene characterisation;
KW intergenomic comparison; chromosome mapping.
XX
OS Eremothecium gossypii.
XX
PN US6239264-B1.
XX
PD 29-MAY-2001.
XX
PF 24-DEC-1997; 97US-00998416.
XX
PR 31-DEC-1996; 97CH-00000016.
XX
PA (SYNG-) SYNGENTA PARTICIPATIONS AG.
XX
PI Philippsen P, Poehlmann R, Steiner-Lange S, Mohr C, Wendland J;
PI Knechtle P, Rebischung C;
XX
DR WPI; 1998-388120/33.
XX
PT New gene for adenylate cyclase from Ashbya gossypii - useful for
PT generating recombinant microorganisms with alteration in gene of cAMP-
PT dependent signalling pathway for increasing production of fine chemicals.
XX
PS Example 3; SEQ ID NO 1035; 632pp; English.
XX
CC The invention relates to isolated DNA molecules comprising isolated
CC genomic DNA sequences from the filamentous funghi Ashbya gossypii, the
CC sequences comprising ADR01309, ADR01366, ADR01367, ADR01388, ADR01428,
CC ADR01466, ADR01629, ADR01637, ADR02057, ADR02345 and ADR02369, chosen
CC from 1047 disclosed genomic sequences. Also included is a cloning vector
CC comprising a nucleotide sequence chosen from the above sequences. The
CC novel Ashbya gossypii genomic sequences are useful for forensic
CC identification, gene characterisation, for studying gene organisation by
CC intergenomic comparison (with Saccharomyces cerevisiae), identifying
CC biosynthetic genes for selectable markers, to isolate

```
CC   promoters/terminators/centromeres, chromosome mapping, and in identifying
CC   sequences unique to Ashbya gossypii for species identification. The
CC   present sequence is an A. gossypii novel genomic sequence of the
CC   invention.
XX
SQ   Sequence 749 BP; 292 A; 84 C; 72 G; 301 T; 0 U; 0 Other;

  Query Match            11.9%;  Score 43.4;  DB 2;  Length 749;
  Best Local Similarity  48.6%;  Pred. No. 0.93;
  Matches  119;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              || || |   | | ||  ||| || || ||||||||    |  || ||||||    ||| |
Db        415 ATAATTAATTAAATAAATAAATAATTAATTAAATTTAAAATGTTTAAAAAAAGAAATAAA 474

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              | |  || | ||  || | |     |||  |  | |||      |  | |  || | |
Db        475 TAATATGTTATATTTAAATAGATCAAAATTTCAACAATTTCCATTTCATTTAGTACTACC 534

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              |     || |  ||    |  |    | | |||  || ||| | |  || |   ||
Db        535 ATCACCATGACCAATTGTTACATCATTTAGTTTATTAGGTTTACTATTAACTTTAGCTTT 594

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
                |||  | |  |  |  | |    | |||   | | ||||| |  | || ||  |  |
Db        595 TACTATACATGGTATTATTGGTAATATTTATCCTTTATTATTATCTTTATTAGTAGTTTT 654

Qy        241 CTGAC 245
                | ||
Db        655 ATTAC 659


RESULT 12
ABL33170
ID   ABL33170 standard; DNA; 15768 BP.
XX
AC   ABL33170;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1143.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1143; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 15768 BP; 3701 A; 401 C; 3826 G; 7840 T; 0 U; 0 Other;

  Query Match            11.9%;  Score 43.4;  DB 6;  Length 15768;
  Best Local Similarity  48.9%;  Pred. No. 1.6;
  Matches  116;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy         29 TTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCTTTAGGTAGAGAACAGGAAA 88
              ||  | ||     ||| |    || || |   |   ||  ||||    |    | | |
Db      15340 TTCGAATTGTTGATTTTAGGTGATTTGTTTATTTCGGTTTTTTAAAGTGTTGGGATTATA 15399

Qy         89 GGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAGTAAAATACTCTCACTAC 148
              ||||   | |  |||  ||   |||  | |    ||| ||  || ||||| | |   |
Db      15400 GGTATGAGTTATTATGTTTAGTTGTAGTATAGGTTTTAATATGTTAAATATTTTTTTTTT 15459

Qy        149 TTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTAGGACAAGAGAGATTTGGTG 208
              |||   | || |  |  | |  ||||  ||   || ||        |||||| ||   |
Db      15460 TTTTATTAATGTGTATGGAAGTTTTAATATTTTTTTTTTATATTTTAGAGAGTTTATATT 15519

Qy        209 TATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACATTTTTGGACTTAGGTATAG 265
              |  | || | || ||  ||| || ||       ||   |||||||  ||   | | |
Db      15520 TTGGAATTAATAATATTAGTTTTTGTTGATAAGTGTTTTTTTTGGGTTTTTTTTTTG 15576


RESULT 13
ABL33527
ID   ABL33527 standard; DNA; 9965 BP.
XX
AC   ABL33527;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1500.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
```

```
PS   Claim 1; SEQ ID NO 1500; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 9965 BP; 2863 A; 187 C; 2139 G; 4776 T; 0 U; 0 Other;

  Query Match             11.7%;  Score 42.8;  DB 6;  Length 9965;
  Best Local Similarity   50.5%;  Pred. No. 2.1;
  Matches  104;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy         10 AATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCT 69
              | |||| |      |  |  || || ||||    |||         |  |  |  ||||  |
Db       2250 ATTAAATTTTTTTATTTGATTTAAGTTTGTTAATTATTTTTGATTTTTTTAAGAAGTTTT 2309

Qy         70 TTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATT 129
                |  |||||    |  ||||||  | ||||| |  || | |   || |    ||||  |
Db       2310 AAAAATAGAGTTTTGATAAGGTATTGTATTATTTAATTTTGTTCGATATTTTATTTGTAT 2369

Qy        130 AGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTA 189
               ||  |||| | |   || | | | |   || || ||||  ||    | ||   ||  ||
Db       2370 TGTTTAATAATTTATTTATTGTTTTTAGGATAAAGTGTAGATTCGTATATATAGTAGGTA 2429

Qy        190 GGACAAGAGAGATTTGGTGTATGGAT 215
              |||       | ||||| | | |  ||
Db       2430 GGATTTATTAAATTTGTTTTGTTTAT 2455


RESULT 14
ACL35363
ID   ACL35363 standard; cDNA; 2000 BP.
XX
AC   ACL35363;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Rice stress-regulated promoter SEQ ID NO:13926.
XX
KW   ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW   agriculture.
XX
OS   Oryza sativa.
XX
PN   WO2003008540-A2.
XX
PD   30-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US019668.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   24-AUG-2001; 2001US-0314662P.
PR   26-SEP-2001; 2001US-0325277P.
PR   21-NOV-2001; 2001US-0332132P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Kreps J,  Briggs SP,  Cooper B,  Glazebrook J,  Goff SA,  Katagiri F;
PI   Moughamer T,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-248011/24.
XX
PT   New stress-responsive nucleic acid, useful for altering the
PT   responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT   stress, salt stress or osmotic stress.
XX
PS   Claim 48; SEQ ID NO 13926; 89pp; English.
XX
CC   The invention relates to novel abiotic stress responsive polynucleotides
CC   and polypeptides. Also disclosed are vectors, expression cassettes, host
CC   cells, and plants containing such polynucleotides. Also disclosed are
CC   methods for using the polynucleotides and polypeptides to alter the
CC   responsiveness of a plant to abiotic stress. The invention is useful in
CC   agriculture. The nucleic acid is useful for determining whether a test
CC   plant has been exposed to an abiotic stress condition. It is also useful
CC   for selecting an agent that alters abiotic stress regulated
CC   polynucleotide expression in a plant cell, and to identify a homolog or
CC   ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC   molecule and the polypeptide encoded by it are useful in altering the
CC   responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC   stress, osmotic stress or any of their combinations. The present sequence
CC   is used in the exemplification of the invention
XX
SQ   Sequence 2000 BP; 540 A; 344 C; 416 G; 428 T; 0 U; 272 Other;

  Query Match             11.6%;  Score 42.6;  DB 11;  Length 2000;
  Best Local Similarity   13.7%;  Pred. No. 1.7;
  Matches   30;  Conservative  105;  Mismatches   84;  Indels    0;  Gaps    0;

Qy          4 ATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCA 63
              :: ::  : |:|:   :: ::::|||:    ::| ::: : ||:::: : ::::|  ::|
Db        109 MWKRWMYMMAWAWSWTRMYWWWKACTYCYTMYKTMYWYRWAAAWRWWKWYKKWWARGWSA 168

Qy         64 AGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGT 123
                 : : :    |  : :::| :::   ::|  ::: ::  : |:      : |:| :: :
Db        169 KMKAYRYRKWCTYYWAWRMAWRWMTTTKWCTWSWYAWWTCYATWMAWATAWTTWTTRMAW 228

Qy        124 TTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCT 183
              :: : :: ::: |:::| : | :: ::: | :| :|: :::  :: :||:::|: : ::
Db        229 WKTRAWMTWWRYAWMYTWWWAKKWAWKYATKWTTYAWRWWMMMWMSRTTYRWMTWAMWYW 288

Qy        184 TATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATT 222
              |: ::   | :    : ::|| : :  : : :::  :|:
Db        289 TWSWYYTWAAMSWYWWCMWTTWSRRRKWYRYKSRKRRTW 327


RESULT 15
ACN44166/c
ID   ACN44166 standard; DNA; 41624 BP.
XX
AC   ACN44166;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG1643838.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 478; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
```

CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 41624 BP; 11435 A; 8284 C; 8267 G; 11480 T; 0 U; 2158 Other;

Query Match 11.6%; Score 42.6; DB 11; Length 41624;
Best Local Similarity 56.9%; Pred. No. 3.1;
Matches 78; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy          6 AATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAG 65
              ||  || ||||    ||||||| || | | ||  |||||  ||||| |    |||||
Db      41510 AAAAAAAAAAAAAGGAAAAATATTTGATTCTTTTACTTTCTAAAAAGAGTTTACATACTT 41451

Qy         66 TGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTT 125
              | ||     ||   |   | ||   || ||  ||| | |  |   |  ||||||  |||
Db      41450 TCCTCCCACTATTTATTTTGTAAACAACTGGCATATTTACCAGATGGGGATTTCATCTTT 41391

Qy        126 GATTAGTAAAATACTCT 142
              |||| ||||  | || |
Db      41390 GATTTGTAATCTGCTTT 41374
```

Search completed: May 10, 2006, 10:59:25
Job time : 172.854 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 11:00:02 ; Search time 1172.23 Seconds
(without alignments)
14608.170 Million cell updates/sec

Title: US-09-423-093B-2_COPY_2744_3109
Perfect score: 366
Sequence: 1 atgataatgaataaaatcaa.....tcactttgctggctggatta 366

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	56.4	15.4	928	10	CNS00DKY	AL071865 Drosophil
c	2	52.4	14.3	1201	10	CNS016FX	AL106695 Drosophil
c	3	51.8	14.2	925	10	CNS006A9	AL064000 Drosophil
	4	50.6	13.8	529	9	B23920	B23920 F17M21TF IG
c	5	50.6	13.8	555	10	AL936843	AL936843 Arabidops
c	6	50.6	13.8	558	10	AL936874	AL936874 Arabidops
c	7	50.6	13.8	1101	10	CNS00LLM	AL068473 Drosophil
	8	49	13.4	467	8	DN385141	DN385141 LIB3892-0
c	9	49	13.4	1101	10	CNS00EO7	AL069440 Drosophil
	10	49	13.4	1201	10	CNS016FX	AL106695 Drosophil
	11	48.2	13.2	1043	10	CNS0145P	AL103735 Drosophil
	12	47.8	13.1	1101	10	CNS0182P	AL108811 Drosophil
c	13	47	12.8	1101	10	CNS00EVL	AL069706 Drosophil
	14	46.8	12.8	534	3	BI303232	BI303232 UI-R-DQ0-
	15	46.8	12.8	928	10	CNS00DKY	AL071865 Drosophil
c	16	46.4	12.7	1488	10	CL461094	CL461094 SAIL_113_
	17	46.2	12.6	1101	10	CNS00YWL	AL096927 Drosophil
c	18	46	12.6	590	9	B59270	B59270 CIT-HSP-201
c	19	46	12.6	615	9	B55701	B55701 CIT-HSP-200
c	20	46	12.6	757	10	AG580281	AG580281 Mus muscu
c	21	45.8	12.5	483	9	BZ437844	BZ437844 BONRM88TF
	22	45.8	12.5	815	10	AG471644	AG471644 Mus muscu
c	23	45.8	12.5	848	9	CC558562	CC558562 CH240_467
	24	45.8	12.5	1225	10	CNS0161D	AL106171 Drosophil
	25	45.4	12.4	877	10	CL099380	CL099380 ISB1-32K7
c	26	45	12.3	267	7	CK573778	CK573778 1875 Plas
	27	44.8	12.2	854	9	BZ428291	BZ428291 BONKG93TR
	28	44.6	12.2	1585	10	AG341261	AG341261 Mus muscu
	29	44.4	12.1	932	9	BH135075	BH135075 ENTOU25TR
	30	44.2	12.1	668	10	CE621964	CE621964 tigr-gss-
	31	44.2	12.1	825	10	CNS009RQ	AL054152 Drosophil
	32	44.2	12.1	1135	10	CL085443	CL085443 ISB1-5A21
c	33	44	12.0	735	9	BZ063475	BZ063475 11f37f01.
	34	44	12.0	817	2	BG782475	BG782475 SEAUMC002
c	35	44	12.0	994	10	CNS0058N	AL057169 Drosophil
c	36	43.8	12.0	529	9	AQ190193	AQ190193 HS_3218_A
	37	43.8	12.0	822	9	BH426886	BH426886 BOGHV24TF
	38	43.8	12.0	872	9	AZ687887	AZ687887 ENTKX16TF
	39	43.8	12.0	1086	10	CNS00YXK	AL096962 Drosophil
c	40	43.6	11.9	629	10	CZ278518	CZ278518 MNTGSS020
c	41	43.6	11.9	794	8	CX069981	CX069981 UCRCS08_1
	42	43.6	11.9	827	10	CNS00EU3	AL069854 Drosophil
	43	43.4	11.9	503	1	AU039139	AU039139 AU039139
c	44	43.4	11.9	689	9	BZ088748	BZ088748 1kf93d04.
	45	43.4	11.9	844	9	CC581314	CC581314 CH240_376

ALIGNMENTS

RESULT 1
CNS00DKY/c
LOCUS CNS00DKY 928 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL071865
VERSION AL071865.1 GI:4948170
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 928)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .928
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR27A24"
/clone_lib="RPCI-98"
/note="end : T7"
ORIGIN

```
  Query Match           15.4%;  Score 56.4;  DB 10;  Length 928;
  Best Local Similarity  30.7%;  Pred. No. 0.014;
  Matches  103;  Conservative   79;  Mismatches  154;  Indels    0;  Gaps    0;

Qy          4 ATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCA 63
              | ||  || ||||  ||||||| |   |||      |    |||||||  |  |::: :|
Db        913 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAWWYTCMA 854

Qy         64 AGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGT 123
              |   :   : : : : ::::  :::  | ::  ::::::::::| | | |: |||    |
Db        853 AAMAMAMMMCVMMMMMMMMMMMMMMMTTMWWHHHWHHHHHHWTTTTTTWTTTTTTTTT 794

Qy        124 TTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCT 183
              ||  ::: ::::::: | : : |: ::: : ::    : :|:  :| ::::  |    :
Db        793 TTTTWWWWWWWWWWWTTTWWWWTWWWWWWWWWWTNNNWNMCHWAWATWWWWWWATATTTW 734

Qy        184 TATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTG 243
              |||:|:  | |: | | |        :|  ||   |:||   : | |       | | |
Db        733 TATWTWAWATAWAAAAAAAAAAAAAAWTAAATTTTTWTTTAWTTWTAAAATTTTTTATT 674

Qy        244 ACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTTGCG 303
              :  ||||:    :|   ||||  |:|:  : :|| |:| |  ||   | ||   |::
Db        673 WTTTTTTWTTTTWTTTWTATATAAWAWTAWAWAAWAWATATTATATWAWATATTTWWTAT 614

Qy        304 TGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGT 339
              |      |||||:: |:  | | :    | |: | |
Db        613 TWAWTATAAAATWWTAWWTTTAAWTAAWATAWAATT 578


RESULT 2
CNS016FX/c
LOCUS       CNS016FX                1201 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15I05 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106695
VERSION     AL106695.1  GI:5623283
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15I05"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match           14.3%;  Score 52.4;  DB 10;  Length 1201;
  Best Local Similarity  33.9%;  Pred. No. 0.12;
  Matches   86;  Conservative   59;  Mismatches  109;  Indels    0;  Gaps    0;

Qy          4 ATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCA 63
              |::|| :|||:|:  ||||||| | |:: :|||:  : |:|:||||:::||   : |: :
Db       1183 AKWATTWATAWAWAWAAAAAAAAAAATWWTWTTTWTWWTTWTWAAAAWWWATATAWWAWGW 1124

Qy         64 AGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGT 123
              :      | :|   |  |  |:   || :   || |:|  || | ::|  ::|    |
Db       1123 WWATAAAAAWAWAAGTATTATTAWTTWTATWWWTTTTTWTWTTTTTTWRTATWWTTTTTT 1064

Qy        124 TTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCT 183
              ||  |:  |  ::|  | |    |  ||| | || | | |: | | : :::   |||| |
Db       1063 TTTTTDTTTTTWWTTTTTTTTKTTTTTTDTATTTTTTRADDTTTTDAKDWTTTTTTAGTT 1004

Qy        184 TATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTG 243
              | |||      :  : :  |||  | | |  ||:: :   :::   ::::: |:: ::: |
Db       1003 TTTTTTTTTKDGTRTRTTTTTTTTTTTTTTATKDKKGRRRMHYMNMHYKDMGMKTHHYTT 944

Qy        244 ACATTTTTGGACTT 257
              | |:  ::|::
Db        943 TTTTMTYNYVMCYK 930


RESULT 3
CNS006A9/c
LOCUS       CNS006A9                925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR13L14 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL064000
VERSION     AL064000.1  GI:4944180
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR13L14"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match           14.2%;  Score 51.8;  DB 10;  Length 925;
  Best Local Similarity  41.1%;  Pred. No. 0.16;
  Matches  109;  Conservative   30;  Mismatches  126;  Indels    0;  Gaps    0;

Qy         13 AAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCTTTA 72
              |||  |  ||:|  : |  ||||| ||    ||:||||:|  |  : ||   |    |   |
Db        922 AAATTATAAWATWAAAATAAATATTTATTTTWAAAAWAKTTWAATKTTTTTTTTTTATAT 863
```

```
Qy          73 GGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAGT 132
               :| : || |  | |  :| | : ||: ||  || | |   ||||    |||  ||  |
Db         862 AWWATWAAAWAAAATATTWAAATDTTTWAATTTTTTTTTTTAAATTTRTTTTTTTTTTTT 803

Qy         133 AAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTAGGA 192
               ||||:  : |   || ||| | || | ||   | |   |||:  |: |   |:|  |  |
Db         802 AAAAWTAWTTATATATTTTATATTTTTTTTTWTTTTTTTTTWTTTKAAWAATWTGAAAAA 743

Qy         193 CAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACATTTTTG 252
               ||| |  |: ::|:| |:        ::|   | | ||      |: | : |  | :||
Db         742 NNAGAAATTTDGKKTKTKTKTTATTTAWKTTDTATTTTTWAAAATKTTTWAAKTTAWTTT 683

Qy         253 GACTTAGGTATAGGAAATGCATTAA 277
                  |:| : | :    ||: |:|||
Db         682 TTTTKAWKAAAWATTGATKAAKTAA 658


RESULT 4
B23920
LOCUS       B23920                 529 bp    DNA     linear   GSS 10-OCT-1997
DEFINITION  F17M21TF IGF Arabidopsis thaliana genomic clone F17M21, genomic
            survey sequence.
ACCESSION   B23920
VERSION     B23920.1  GI:2509551
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Coffey,E.,
            Golden,K., Johnson,K., Adams,M.D. and Venter,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: F17M21TR
            Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 529.
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="F17M21"
                     /sex="hermaphrodite"
                     /clone_lib="IGF"
                     /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
                     Produced by Thomas Altmann"
ORIGIN

  Query Match             13.8%;  Score 50.6;  DB 9;  Length 529;
  Best Local Similarity   46.9%;  Pred. No. 0.3;
  Matches  158;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy           1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
               || | ||  ||  |||   |||   | | || || |  ||  |  | || |||||    |
Db          34 ATAAGAACAAAATAAACAGAAATCGTTCATACATATCCCATATCGATAAGATATGTAGTA 93

Qy          61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
               |||||   |||  || | || |  |||| |||| ||   ||   |   |  |||| ||
Db          94 TCAAGATACTTAATTAAAAAAAATAAAAGTTACATAATCGGATAAGTCAATACTATTGCA 153

Qy         121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
```

```
               | ||  |   ||  || | | |   |||||||| | |   | ||   |  |  |
Db         154 ATTTACAAATATATGTTAGTTTAATATTTTTCTCTTAGAATCTATATATACATTTTTGCA 213

Qy         181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
               |  |||  || | |    | | ||        |   ||  |   |    ||   |
Db         214 GTTATTTTGTGAAATAAATTCTGTAGTTGGGACTTATTTACAATGGCTGCCACTGAATTT 273

Qy         241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
               || ||     ||  || | |||    ||   ||  |  ||   ||||| |
Db         274 TAATGTTTGTTTTTAATAATTAAAAAAAACATCTTCGAAATAAATATTTGACATTTAACA 333

Qy         301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTA 337
                 | |   |||| | |  | || |||    ||| ||
Db         334 ATCTTTCCTAAAACTCTCCACATTAGTTACTAAAATA 370


RESULT 5
AL936843/c
LOCUS       AL936843               555 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-061D12-016089,
            genomic survey sequence.
ACCESSION   AL936843
VERSION     AL936843.1  GI:24368468
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
  TITLE     GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for
            the identification of T-DNA insertion mutants in Arabidopsis
            thaliana
  JOURNAL   Bioinformatics 19 (11), 1441-1442 (2003)
   PUBMED   12874060
REFERENCE   2
  AUTHORS   Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and
            Weisshaar,B.
  TITLE     An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for
            flanking sequence tag-based reverse genetics
  JOURNAL   Plant Mol. Biol. 53 (1-2), 247-259 (2003)
   PUBMED   14756321
REFERENCE   3
  AUTHORS   Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and
            Weisshaar,B.
  TITLE     High-throughput generation of sequence indexes from T-DNA
            mutagenized Arabidopsis thaliana lines
  JOURNAL   BioTechniques 35 (6), 1164-1168 (2003)
   PUBMED   14682050
REFERENCE   4  (bases 1 to 555)
  AUTHORS   Strizhov,N., Li,Y., Rosso,M.G. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence has been recovered from the left border of the T-DNA.
            It indicates an insertion within the locus defined by BAC clone
            K13P22. Details on the protocols used for generation of the
            sequence are described in References 1-3. The sequences are
            generated at the MPI for Plant Breeding Research in the context of
            the GABI-Kat project. GABI-Kat is part of the German Plant Genomics
            program designated 'GABI'. Information on line availability can be
            found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="GK-061D12-016089"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Col-0"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
```

vector pAC161 (GenBank accession number: AJ537514). The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed."

ORIGIN

Query Match 13.8%; Score 50.6; DB 10; Length 555;
Best Local Similarity 46.9%; Pred. No. 0.3;
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              || | ||  ||  |||   |||   | | || || |  ||  |  | || |||||    |
Db        467 ATAAGAACAAAATAAACAGAAATCGTTCATACATATCCCATATCGATAAGATATGTAGTA 408

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              |||||    |||  || | || |  |||| ||||  ||   ||   |   |  |||| ||
Db        407 TCAAGATACTTAATTAAAAAAAATAAAAGTTACATAATCGGATAAGTCAATACTATTGCA 348

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              | ||  |    ||  || | | |   |||||||| | |   | ||   |   |  |
Db        347 ATTTACAAATATATGTTAGTTTAATATTTTTCTCTTAGAATCTATATATACATTTTTGCA 288

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              |  |||  || |  |    | | ||     |   ||  |    |      ||    |
Db        287 GTTATTTTGTGAAATAAATTCTGTAGTTGGGACTTATTTACAATGGCTGCCACTGAATTT 228

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              || ||     ||  || |  |||    ||   ||  |  ||   |||||  |
Db        227 TAATGTTTGTTTTTAATAATTAAAAAAAACATCTTCGAAATAAATATTTGACATTTAACA 168

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTA 337
                | |   |||| | |  | || |||    ||| ||
Db        167 ATCTTTCCTAAAACTCTCCACATTAGTTACTAAAATA 131
```

RESULT 6
AL936874/c
LOCUS AL936874 558 bp DNA linear GSS 01-APR-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-061F12-016089, genomic survey sequence.
ACCESSION AL936874
VERSION AL936874.1 GI:24368499
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
TITLE GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in Arabidopsis thaliana
JOURNAL Bioinformatics 19 (11), 1441-1442 (2003)
PUBMED 12874060
REFERENCE 2
AUTHORS Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and Weisshaar,B.
TITLE An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics
JOURNAL Plant Mol. Biol. 53 (1-2), 247-259 (2003)
PUBMED 14756321
REFERENCE 3
AUTHORS Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and Weisshaar,B.
TITLE High-throughput generation of sequence indexes from T-DNA mutagenized Arabidopsis thaliana lines
JOURNAL BioTechniques 35 (6), 1164-1168 (2003)
PUBMED 14682050
REFERENCE 4 (bases 1 to 558)
AUTHORS Rosso,M.G., Strizhov,N., Li,Y. and Weisshaar,B.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence has been recovered from the left border of the T-DNA. It indicates an insertion within the locus defined by BAC clone K13P22. Details on the protocols used for generation of the sequence are described in References 1-3. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES Location/Qualifiers
source 1. .558
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="GK-061F12-016089"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Col-0"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161 (GenBank accession number: AJ537514). The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed."
ORIGIN

Query Match 13.8%; Score 50.6; DB 10; Length 558;
Best Local Similarity 46.9%; Pred. No. 0.3;
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              || | ||  ||  |||   |||   | | || || |  ||  |  | || |||||    |
Db        468 ATAAGAACAAAATAAACAGAAATCGTTCATACATATCCCATATCGATAAGATATGTAGTA 409

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              |||||    |||  || | || |  |||| ||||  ||   ||   |   |  |||| ||
Db        408 TCAAGATACTTAATTAAAAAAAATAAAAGTTACATAATCGGATAAGTCAATACTATTGCA 349

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              | ||  |    ||  || | | |   |||||||| | |   | ||   |   |  |
Db        348 ATTTACAAATATATGTTAGTTTAATATTTTTCTCTTAGAATCTATATATACATTTTTGCA 289

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              |  |||  || |  |    | | ||     |   ||  |    |      ||    |
Db        288 GTTATTTTGTGAAATAAATTCTGTAGTTGGGACTTATTTACAATGGCTGCCACTGAATTT 229

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              || ||     ||  || |  |||    ||   ||  |  ||   |||||  |
Db        228 TAATGTTTGTTTTTAATAATTAAAAAAAACATCTTCGAAATAAATATTTGACATTTAACA 169

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTA 337
                | |   |||| | |  | || |||    ||| ||
Db        168 ATCTTTCCTAAAACTCTCCACATTAGTTACTAAAATA 132
```

RESULT 7
CNS00LLM/c
LOCUS CNS00LLM 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR30I18 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL068473
VERSION AL068473.1 GI:4958603
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission

```
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR30I18"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             13.8%;  Score 50.6;  DB 10;  Length 1101;
  Best Local Similarity   28.4%;  Pred. No. 0.31;
  Matches   85;  Conservative   81;  Mismatches  133;  Indels    0;  Gaps    0;

Qy         10 AATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCT 69
              :| :||    |:|:| ||  :|:|  ::  : |  :| |  |:|:  |: |  ||:|  |
Db        835 DAAWAATRAGARARAGTADNWADAANKWTARATNGWAKATGADAKATTRTAAAAADTTTT 776

Qy         70 TTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATT 129
              :|  ||    :: |:  :: :  |: :::::::  :: : |:|: | | | ||:|:::
Db        775 WTTWGTTAKAWWTAKAGNRKAWTDADTRKKWKWKARWKADTTRTDTTGTTATGTKTKWKA 716

Qy        130 AGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTA 189
              :   | |||   : :    |||    | |::  |  :  |  |||  |:    |    |:
Db        715 DTNKAGATAAGTWADARTATTTNRAATGTDKAKATAKAKATATTATDTKAKGKTATGATR 656

Qy        190 GGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACATTT 249
              :  |||||:|:|  :||||::  : | ||:|:  ::| ||:  : :  : : : | |
Db        655 TAWGKAGAGARADTGTDTGTADRWGWTATTADTWTTDRTWTTKWAKNKTAKAKAWGAGTN 596

Qy        250 TTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTTGCGTGTGG 308
              : : :     || :|::  :|: |: :  :|: ||: :   :   : :|  : || |||
Db        595 KARTRKAAGTGTGWAKRKKDTRTADGRGGDARTAGRGDGAGDTTGWGRTRGKAGTTTGG 537


RESULT 8
DN385141
LOCUS       DN385141                467 bp    mRNA    linear   EST 07-MAR-2005
DEFINITION  LIB3892-008-Q6-K1-C11 LIB3892 Canis familiaris cDNA clone
            CLN8888666, mRNA sequence.
ACCESSION   DN385141
VERSION     DN385141.1  GI:60566362
KEYWORDS    EST.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Staten,N.R.
  TITLE     Direct Submission (Staten,N.R.)
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Nick Staten
            Tel: 636 247 6855
            Email: nicholas.r.staten@pfizer.com.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Canis familiaris"
                     /mol_type="mRNA"
                     /strain="beagle"
                     /db_xref="taxon:9615"
                     /clone="CLN8888666"
                     /tissue_type="heart"
                     /lab_host="BHK"
                     /clone_lib="LIB3892"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI;
                     K9-left ventricle"
ORIGIN

  Query Match             13.4%;  Score 49;  DB 8;  Length 467;
  Best Local Similarity   52.7%;  Pred. No. 0.7;
  Matches  106;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy          4 ATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCA 63
              | ||  || ||||  ||  |||||  ||||  ||  |    |||||||  |  ||| |  |
Db          9 AAAAAAAAAAAAAAAAACTAAATATATAAAAATTTTAAAAAAAAAAAAAAAAAATAAAAAA 68

Qy         64 AGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGT 123
              ||      |    | | ||  || | |||    |   ||| |  || |    ||||| ||
Db         69 AGAAAAATTATAAAAAAAGGGGTAGGGTGTTGATATATTTAATTATTAAATATTTAAACA 128

Qy        124 TTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCT 183
                | | | ||| |   | |||  ||| |  | | ||||     |   | | ||| | |
Db        129 AGTTTAATTTAAAAATAATAACTTTTTTTTTATTTTTTAATAAATAAAATTAATTTTCTT 188

Qy        184 TATTTAGGACAAGAGAGATTT 204
              | |||   | |||  | ||||
Db        189 TCTTTTTTAAAAGTTAAATTT 209


RESULT 9
CNS00EO7/c
LOCUS       CNS00EO7                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069440
VERSION     AL069440.1  GI:4949583
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29P01"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             13.4%;  Score 49;  DB 10;  Length 1101;
  Best Local Similarity   37.0%;  Pred. No. 0.71;
  Matches  100;  Conservative   42;  Mismatches  128;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||  :||   | |||:| | ::|| |  : ||:  ||    || |:: || | :  ||||
Db        936 ATTTWAAAATAAAAAWTAACWWAATTTAWAAAWCATTWTTTTTAAWWTAATTTWATTACA 877

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              : :     |  |   | :  | |  |||   | |: ||||:|   |    : |  :|  |
Db        876 WTWTTAWWTAAAAAAATWTTAAAWTAAAWAAAAARWATTAWAATTTAYATWATATWTAAA 817

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              : | | :  | ::: | : |   :  ||| | ||:||: |: : |:| : : | |||
Db        816 WWTATAWATTATTMWWAATWTTATAWNATTTTTTTTWTAWAAWTWTTWATWAWTAATTTT 757

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              :|::|||    || | |  ||   | |||  :| | |:  |   | ||  :  |  :
Db        756 AAWTWWTTAAATAAAAAAAATTTATTTTTATTTWTTATTWAAAWTTTTWTTTTWAATTWW 697

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAAT 270
              :|   || :||  | |:    |:|   ||:
Db        696 YTTTAATAWTTAAAWTWTTAAAWAWWTAAW 667


RESULT 10
CNS016FX
LOCUS       CNS016FX                1201 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15I05 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106695
VERSION     AL106695.1  GI:5623283
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15I05"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             13.4%;  Score 49;  DB 10;  Length 1201;
  Best Local Similarity   41.4%;  Pred. No. 0.71;
  Matches   75;  Conservative   33;  Mismatches   73;  Indels    0;  Gaps    0;

Qy          4 ATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCA 63
              | |::  | ||||| :|||||| |   ||| ::   | :  ||||||| |: |||::  |
Db       1020 AAAHHTYAAAAAATAHAAAAAAMAAAAAAAAWWAAAAAHAAAAAAAAAAAWWATAYWAAA 1079

Qy         64 AGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGT 123
              |    :  |  :: | |: |  ||   |||    :| | |  ::   | :||| | :: |
Db       1080 AAAWAWAAAAAWWWATAWAAWTAATAATACTTWTWTTTTTATWWWCWTWWTATATWWWTT 1139

Qy        124 TTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCT 183
              || |: | :|:|| : :    |  ||| : :|:|||:|| : ||  ||::  |||:  :
Db       1140 TTWAWAAWWAWAAAWAWWATTTTTTTTTTWTWTWTATWAATWMTATTTTWWTTTTTWTAH 1199

Qy        184 T 184
              |
Db       1200 T 1200


RESULT 11
CNS0145P
LOCUS       CNS0145P                1043 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN11G11 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL103735
VERSION     AL103735.1  GI:5615346
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1043)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1043
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN11G11"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             13.2%;  Score 48.2;  DB 10;  Length 1043;
  Best Local Similarity   36.6%;  Pred. No. 1.1;
  Matches  100;  Conservative   43;  Mismatches  130;  Indels    0;  Gaps    0;

Qy         11 ATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCTT 70
              ||| | : ||:|| ||:  || :|||    ::|: :   ||:| ::| ||  | :  ||
Db        600 ATATATWAAAWAATATWTATATWTTTTTTTWWWTWTWTTTATWTWWWAAATATATWTTTT 659

Qy         71 TAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTA 130
              :   :      :   |   || :  :|: :||   | | | ||||:|  | ||  ||:
Db        660 WTWTWTWTTTTTWTWTATATTATWWTWTWTWATAAATATTTTTTATWTATATTTATATWT 719

Qy        131 GTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTAG 190
              |:   :| | |   || ||:   |  | ||   | |   |||   |||   || |||
Db        720 ATWTTTWAATTTATATATTTWANATANTTTTTTTTTTTTTTTTTTTTTTTTTTTTTA 779
```

```
Qy        191 GACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACATTTT 250
              | || | | :::: : :|  :: : :|: |  | | ||  | ::  :|    |: |
Db        780 AAAAAWAAAAWWWWWWWWWAAWWWWWWWWAWAAWAATTTTTTTTTHHTTYCCYYTAWATA 839

Qy        251 TGGACTTAGGTATAGGAAATGCATTAACAAACA 283
              |  :  :|  :|:|  |||   |  || ||| |
Db        840 TATWWAWATAWAWAWAAAAAAAAAAAAAAAAAA 872


RESULT 12
CNS0182P
LOCUS       CNS0182P               1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108811
VERSION     AL108811.1  GI:5629115
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN37D10"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             13.1%;  Score 47.8;  DB 10;  Length 1101;
  Best Local Similarity   14.9%;  Pred. No. 1.3;
  Matches   41;  Conservative  142;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              | :| || :||:::::: :|||||| |    ||:||: : : :::::: :::::| | : :
Db        812 ARRARAARRAADDRDWDAWAAAAAAAAAAAAAAAWTTWRDRWWDWDDWDTRWDDDTTAAWWD 871

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              : :|: :    ::: ::: :: |::||:: :  :: |:  |::  |: : |:::::::
Db        872 DARARARRRRRRRRRRRRARRAARRAADDTDTKDRWADATTDKDTTKTWTTDDDDWDKAK 931

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              :: :: |: :| : ::::   : | || ::: : :::     :  ::  ::::    :::
Db        932 RDRWWAAKADGAWKWRDRARDDWAATAKDDDGWKDKWGGRGKRGGKDKKRWDKGKTGKKD 991

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
                ::::|: ::: ::  : ::::: ::|   |::|:  :  ::  : :  :: :| :  |
Db        992 DDDWDKTWTRDWWWWWTRTKWDWWWDDGRGRGRWTRKRWGAWWRADAWARDDTDGKDTRT 1051

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATT 275
               :::   |:||::: ::::::::::::::: :::
Db       1052 ADKDRKRTDTTKRDGDDWRKDRKDKKRRDKKGDDK 1086
```

```
RESULT 13
CNS00EVL/c
LOCUS       CNS00EVL               1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             12.8%;  Score 47;  DB 10;  Length 1101;
  Best Local Similarity   34.5%;  Pred. No. 2;
  Matches  117;  Conservative   62;  Mismatches  157;  Indels    3;  Gaps    1;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||  ||:  |   ||:| |  :: :||  |: |:::|  : | ||  |:  |:| | | |
Db       1079 ATWTTAWATATATAAWTAATTWWTWTATATWTAWWWTATWATATATTAWTTTWTAAAATA 1020

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              | :: |  |  |  || : |: |  | |  || |  :  :|||   : :   |:| |: |
Db       1019 TAWWTTAATAAATATATWTAWTAWWATATATAWATAWAAWTATAWAWTWWAATWAATWWA 960

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
                |  |:|   | ||:| | | | :: :|| | || | ||    | |   | |:  ||||
Db        959 ---TAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTTTTTWTTTWTATWTWTTTA 903

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              |::|::|  | |: | : :| |  | |:|    :   :||    : | ||  :  |  |
Db        902 TATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATATTAWATTTTTTWTWTATT 843

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              : | ||: |:  | :|:  | |:  :  ||  ||| | : : :  :|   : ||: || :
Db        842 WWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTWTWTTWTATTWTATWTATAW 783

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGT 339
                :     : |:| :|| | || : |     ::|:| |
Db        782 WWTWWATWWWTAWAAWTATATATWWATATTATWWTWATT 744
```

```
RESULT 14
BI303232
LOCUS       BI303232                534 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  UI-R-DQ0-ciz-b-17-0-UI.s1 UI-R-DQ0 Rattus norvegicus cDNA clone
            UI-R-DQ0-ciz-b-17-0-UI 3', mRNA sequence.
ACCESSION   BI303232
VERSION     BI303232.1  GI:14979512
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 534)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized rat cell line R3327-5a library cDNA Library
            Preparation: M.B. Soares Lab Clone distribution: clones will be
            available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .534
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-DQ0-ciz-b-17-0-UI"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-DQ0"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DQ0
                     library is a non-normalized Rat Cell Line R3327-5A library
                     constructed in pT3T7 PAC vector according to the procedure
                     described by Bonaldo, Lennon & Soares (Genome Research
                     Genome 6: 791-806, 1996). The oligonucleotide used to
                     prime first strand synthesis contained the sequence tag
                     CCAGG between the Not I cloning site and dT18 stretch.
                     TAG_TISSUE=rat cell line R3327-5a
                     TAG_LIB=UI-R-DQ0
                     TAG_SEQ=GGACTAGATC"
ORIGIN
```

```
  Query Match           12.8%;  Score 46.8;  DB 3;  Length 534;
  Best Local Similarity   46.8%;  Pred. No. 2.2;
  Matches  147;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy          2 TGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACAT 61
              ||| ||  | | ||||  | | | |       |  |||| | | | |||| || || |
Db        147 TGACAAACATTTAAATTGACAGACTCAAGTTGATAGGCACATAATACAAATTTGGTAAAA 206

Qy         62 CAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAA 121
              |  ||     |||  | |     | |     | |  |    |    ||   |
Db        207 CGTGTCTCAGAGGCTAACACTGAAGCACATCTGTTTTCAAGACTCCATAAAAAATCCAGA 266

Qy        122 GTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTAC 181
              ||  ||  ||||    |  | | ||| |||||  |    ||  ||  | |
Db        267 CTTCACTTGCCAAAAAGTCCAATCAATTTTGTCTTAGCATGTAACTACAAATAGATAAAA 326

Qy        182 CTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTC 241
              |||| | | |||  |  |||||| |    |||  |||    | || |   |  |
Db        327 TAAATTTTGCATAAGTCAATCTTGGTGTTTTATGGACATTTAATTTTTTTAGCATTAAGC 386

Qy        242 TGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTTG 301
              ||  ||| ||  |    | | || |   | |||| |||||  |||||     | |||||
Db        387 TGTGATTATTTTAAGAGGCTGTATGTCCTCCATTCACAAAAGGGATCAACATTACATTTA 446

Qy        302 CGTGTGGCAAAAAT 315
              | |   |  |||||
Db        447 CTTTAAGTTAAAAT 460
```

```
RESULT 15
CNS00DKY
LOCUS       CNS00DKY                928 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR27A24"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN
```

```
  Query Match           12.8%;  Score 46.8;  DB 10;  Length 928;
  Best Local Similarity   27.6%;  Pred. No. 2.3;
  Matches   68;  Conservative   73;  Mismatches  105;  Indels    0;  Gaps    0;

Qy          6 AATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAG 65
              |||  :|:| || | | : :|  |: |:|:||  |  : ||:||||:|  | | |  ||
Db        620 AATATWTWATATAATATWTWTTWTWTAWTWTTATATAWAAAWAAAAWAAAAAAWAATAAA 679

Qy         66 TGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTT 125
              |||    |   || :  |||  ||     || | |  || | | |:|: | :|  |
Db        680 AAATTTTAWAAWTAAAAWAAAAATTTAWTTTTTTTTTTTTTTWTATWTWAWATAWAAATA 739
```

```
Qy        126 GATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTA 185
              :::: :| : :: : :    : ::: : ::|::::|| : ::: :::::    |    |
Db        740 TWWWWWWATWTWDGKNWNNNAWWWWWWWWWWAWWWWAAAWWWWWWWWWWAAAAAAAAAA 799

Qy        186 TTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGAC 245
                :  | || :::::::::::: :   : ::  : ::    ::: :::|: :  |   :
Db        800 AAAWAAAAAAAAWDDDDDDWDDDWWKAAKKKKKKKKKKKKKKKKKBGKKKTKTKTTTKGARW 859

Qy        246 ATTTTT 251
              :|||||
Db        860 WTTTTT 865
```

Search completed: May 11, 2006, 05:00:40
Job time : 1175.23 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 20:34:42 ; Search time 51.0518 Seconds
(without alignments)
12743.667 Million cell updates/sec

Title: US-09-423-093B-2_COPY_2744_3109
Perfect score: 366
Sequence: 1 atgataatgaataaaatcaa.....tcactttgctggctggatta 366

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	366	100.0	14187	3	US-09-453-702B-121	Sequence 121, App
c	2	366	100.0	14187	3	US-10-114-170-121	Sequence 121, App
c	3	43.4	11.9	601	3	US-09-949-016-149640	Sequence 149640,
	4	43.4	11.9	728	3	US-08-998-416-702	Sequence 702, App
	5	43.4	11.9	732	3	US-08-998-416-938	Sequence 938, App
	6	43.4	11.9	749	3	US-08-998-416-1035	Sequence 1035, Ap
c	7	42.2	11.5	140725	3	US-09-949-016-17074	Sequence 17074, A
c	8	42.2	11.5	149543	3	US-09-949-016-15947	Sequence 15947, A
c	9	41.8	11.4	601	3	US-09-949-016-149641	Sequence 149641,
	10	41.8	11.4	708	3	US-08-998-416-785	Sequence 785, App
c	11	40.8	11.1	210	3	US-09-248-796A-13125	Sequence 13125, A
c	12	40.8	11.1	601	3	US-09-949-016-196365	Sequence 196365,
c	13	40.8	11.1	601	3	US-09-949-016-196366	Sequence 196366,
c	14	40.8	11.1	601	3	US-09-949-016-196367	Sequence 196367,
c	15	40.8	11.1	451924	3	US-09-949-016-12896	Sequence 12896, A
c	16	40.8	11.1	451925	3	US-09-949-016-17305	Sequence 17305, A
c	17	40.6	11.1	134987	3	US-09-949-016-15348	Sequence 15348, A
c	18	40.6	11.1	134987	3	US-09-949-016-15349	Sequence 15349, A
c	19	40.6	11.1	134987	3	US-09-949-016-15350	Sequence 15350, A
c	20	40.6	11.1	134987	3	US-09-949-016-15507	Sequence 15507, A
c	21	40.6	11.1	134987	3	US-09-949-016-15508	Sequence 15508, A
c	22	40.6	11.1	134987	3	US-09-949-016-15509	Sequence 15509, A
	23	39.6	10.8	119153	3	US-09-949-016-12378	Sequence 12378, A
	24	39.4	10.8	641	3	US-08-998-416-184	Sequence 184, App
	25	39	10.7	88245	3	US-09-949-016-13835	Sequence 13835, A
c	26	38.8	10.6	58014	3	US-09-949-016-17448	Sequence 17448, A
	27	38.6	10.5	601	3	US-09-949-016-70808	Sequence 70808, A
c	28	38.6	10.5	1926	3	US-09-583-110-1602	Sequence 1602, Ap
c	29	38.6	10.5	1926	3	US-09-107-433-345	Sequence 345, App
c	30	38.6	10.5	225127	3	US-09-949-016-16480	Sequence 16480, A
c	31	38.2	10.4	601	3	US-09-949-016-168196	Sequence 168196,
c	32	38.2	10.4	601	3	US-09-949-002-7548	Sequence 7548, Ap
	33	38.2	10.4	25590	3	US-09-949-002-777	Sequence 777, App
c	34	37.8	10.3	601	3	US-09-949-002-7549	Sequence 7549, Ap
c	35	37.8	10.3	601	3	US-09-949-002-7550	Sequence 7550, Ap
	36	37.8	10.3	2216	3	US-09-426-783-5	Sequence 5, Appli
c	37	37.8	10.3	2621	2	US-08-553-619B-8	Sequence 8, Appli
c	38	37.8	10.3	72604	3	US-09-268-992-7	Sequence 7, Appli
c	39	37.8	10.3	72604	3	US-09-657-474-7	Sequence 7, Appli
	40	37.6	10.3	2434	3	US-09-489-847-67	Sequence 67, Appl
c	41	37.4	10.2	1767	3	US-09-248-796A-6192	Sequence 6192, Ap
	42	37.2	10.2	26191	3	US-09-949-002-686	Sequence 686, App
	43	37.2	10.2	39428	3	US-09-949-002-794	Sequence 794, App
c	44	37.2	10.2	116652	3	US-09-949-016-13413	Sequence 13413, A
	45	37	10.1	41696	3	US-09-949-016-12444	Sequence 12444, A

ALIGNMENTS

RESULT 1
US-09-453-702B-121/c
; Sequence 121, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14187
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-09-453-702B-121

```
  Query Match           100.0%;  Score 366;  DB 3;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 6.8e-86;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11257 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 11198

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11197 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 11138

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11137 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 11078

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11077 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 11018

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11017 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 10958

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10957 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 10898

Qy        361 GGATTA 366
              ||||||
Db      10897 GGATTA 10892
```

RESULT 2
US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Patent No. 6855814
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6855814el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14187
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121


```
  Query Match           100.0%;  Score 366;  DB 3;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 6.8e-86;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11257 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 11198

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11197 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 11138

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11137 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 11078

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11077 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 11018

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11017 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 10958

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10957 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 10898

Qy        361 GGATTA 366
              ||||||
Db      10897 GGATTA 10892
```


RESULT 3
US-09-949-016-149640/c
; Sequence 149640, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149640
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-149640

```
  Query Match           11.9%;  Score 43.4;  DB 3;  Length 601;
  Best Local Similarity  58.3%;  Pred. No. 0.05;
  Matches   74;  Conservative    1;  Mismatches   52;  Indels    0;  Gaps    0;
```

```
Qy          14 AAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCTTTAG 73
               |||  ||||||  |:   || |    ||   ||| |||  | | || ||||  | ||||
Db         315 AAAAAAAAAAATAAYAGTAAGTAAAAACAAAAAACAAAATTAAGACTTCAATAGGTTTAA 256

Qy          74 GTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAGTA 133
               | | | |||||||||   ||| ||  | |  ||| |     | ||||||||  || |
Db         255 GAAAACAACAGGAAAAAAACAAGAACAAAGGCTTTTAATGGAGTTCAAGTTCCATAAAAC 196

Qy         134 AAATACT 140
               | ||| |
Db         195 ACATATT 189


RESULT 4
US-08-998-416-702
; Sequence 702, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  702:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  728 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1462UP
US-08-998-416-702

  Query Match            11.9%;  Score 43.4;  DB 3;  Length 728;
  Best Local Similarity  48.6%;  Pred. No. 0.053;
  Matches  119;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy           1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
               || || |   | | ||  ||| || || ||||||||    |  || ||||||    ||| |
Db         415 ATAATTAATTAAATAAATAAATAATTAATTAAATTTAAAATGTTTAAAAAAAGAAATAAA 474

Qy          61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
               | |  || | ||  || | |     ||| |  | |||       |  | |  || | |
Db         475 TAATATGTTATATTTAAATAGATCAAAATTTCAACAATTTCCATTTCATTTAGTACTACC 534

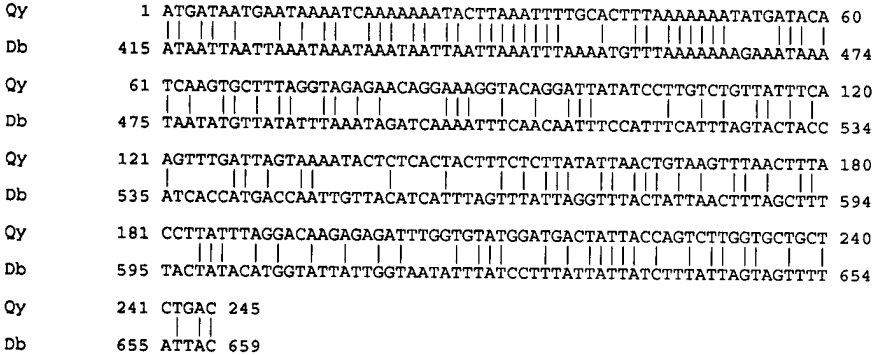
Qy         121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
               |      || | ||      |    |   | | ||| || ||| | |  || |  ||
Db         535 ATCACCATGACCAATTGTTACATCATTTAGTTTATTAGGTTTACTATTAACTTTAGCTTT 594

Qy         181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
               |||  | |  |  |  | |    | |||   | | ||||| |  | || ||  |  |
Db         595 TACTATACATGGTATTATTGGTAATATTTATCCTTTATTATTATCTTTATTAGTAGTTTT 654

Qy         241 CTGAC 245
               | ||
Db         655 ATTAC 659

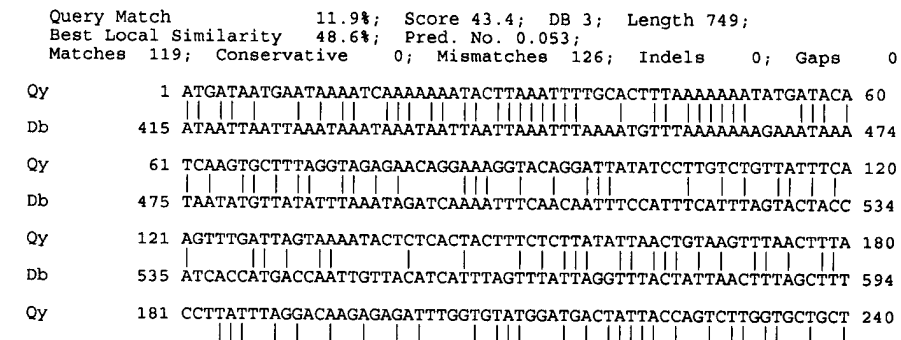

RESULT 5
US-08-998-416-938
; Sequence 938, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  938:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  732 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1581UP
US-08-998-416-938

  Query Match            11.9%;  Score 43.4;  DB 3;  Length 732;
  Best Local Similarity  48.6%;  Pred. No. 0.053;
  Matches  119;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;
```



RESULT 6
US-08-998-416-1035
; Sequence 1035, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1035:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 749 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1633RP
US-08-998-416-1035

Query Match 11.9%; Score 43.4; DB 3; Length 749;
Best Local Similarity 48.6%; Pred. No. 0.053;
Matches 119; Conservative 0; Mismatches 126; Indels 0; Gaps 0;



RESULT 7
US-09-949-016-17074/c
; Sequence 17074, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17074
; LENGTH: 140725
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17074

Query Match 11.5%; Score 42.2; DB 3; Length 140725;
Best Local Similarity 51.2%; Pred. No. 0.42;
Matches 154; Conservative 0; Mismatches 138; Indels 9; Gaps 2;



```
Qy        301 G 301
              |
Db      73018 G 73018
```

RESULT 8
US-09-949-016-15947/c
; Sequence 15947, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15947
; LENGTH: 149543
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15947

Query Match 11.5%; Score 42.2; DB 3; Length 149543;
Best Local Similarity 58.3%; Pred. No. 0.43;
Matches 74; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy         14 AAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCTTTAG 73
              |||  ||||||  |    || |    ||   ||| |||  | | || ||||  | ||||
Db       6621 AAAAAAAAAAATAATAGTAAGTAAAAACAAAAAACAAAATTAAGACTTCAATAGGTTTAA 6562

Qy         74 GTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAGTA 133
              | | | |||||||||   ||| ||  | |  ||| |     | ||||||||  || |
Db       6561 GAAAACAACAGGAAAAAAACAAGAACAAAGGCTTTTAATGGAGTTCAAGTTCCATAAAAC 6502

Qy        134 AAATACT 140
              | ||| |
Db       6501 ACATATT 6495
```

RESULT 9
US-09-949-016-149641/c
; Sequence 149641, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149641
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-149641

Query Match 11.4%; Score 41.8; DB 3; Length 601;
Best Local Similarity 57.5%; Pred. No. 0.13;
Matches 73; Conservative 1; Mismatches 53; Indels 0; Gaps 0;

```
Qy         14 AAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCTTTAG 73
              :||  ||||||  |    || |    ||   ||| |||  | | || ||||  | ||||
Db        301 MAAAAAAAAAATAATAGTAAGTAAAAACAAAAAACAAAATTAAGACTTCAATAGGTTTAA 242

Qy         74 GTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAGTA 133
              | | | |||||||||   ||| ||  | |  ||| |     | ||||||||  || |
Db        241 GAAAACAACAGGAAAAAAACAAGAACAAAGGCTTTTAATGGAGTTCAAGTTCCATAAAAC 182

Qy        134 AAATACT 140
              | ||| |
Db        181 ACATATT 175
```

RESULT 10
US-08-998-416-785
; Sequence 785, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 785:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 708 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1504RP
US-08-998-416-785

Query Match 11.4%; Score 41.8; DB 3; Length 708;
Best Local Similarity 48.2%; Pred. No. 0.14;
Matches 118; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```
Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              || ||  |   | | ||  ||| || || |||||||||   | || ||||||    ||| |
Db        415 ATAATTAATTAAATAAATAAATAATTAATTAAATTTAAAATGTTTAAAAAAAGAAATAAA 474

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              |  |  ||   ||  || | |     |||  |  |  |||       |  |  || | |
Db        475 TAATATGTGATATTTAAATAGATCAAAATTTCAACAATTTCCATTTCATTTAGTACTACC 534

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              |      ||  ||        |    |  | | |||  || ||| | |  || |  ||
Db        535 ATCACCATGACCAATTGTTACATCATTTAGTTTATTAGGTTTACTATTAACTTTAGCTTT 594

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              |||  | |  |  | |  | |    | |||   | |||||  |  | || ||  | |
Db        595 TACTATACATGGTATTATTGGTAATATTTATCCCTTATTATTATCTTTATTAGTAGTTTT 654

Qy        241 CTGAC 245
              | ||
Db        655 ATTAC 659
```

RESULT 11
US-09-248-796A-13125/c
; Sequence 13125, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 13125
; LENGTH: 210
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-13125

Query Match 11.1%; Score 40.8; DB 3; Length 210;
Best Local Similarity 52.3%; Pred. No. 0.18;
Matches 90; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

```
Qy         80 AACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAGTAAAATAC 139
              ||||  || |  || |||| | |   | |     ||| | ||| |||    | || |
Db        181 AACAAAAATGACACGGGATAAATTTTATTTTATAAATTCTAGGTTCGATGCTGATAAGAA 122

Qy        140 TCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTAGGACAAGAGA 199
              |    | || ||   ||  | |||  | |   ||| |||||   || || |    ||  |
Db        121 TGCAATTAATTATGCTATTTTTACTTTTTTTTTTCACTTTTTTTTCTTGAATGTAAAGAA 62

Qy        200 GATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACATTTTT 251
              ||  | ||||  | |||  |||||    ||   |  |||  | || ||||
Db         61 TTATTAATCTATGTGTAACTGTTACCTTAATTCTAGAAGCTGAGTCACTTTT 10
```

RESULT 12
US-09-949-016-196365/c
; Sequence 196365, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 196365
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-196365

Query Match 11.1%; Score 40.8; DB 3; Length 601;
Best Local Similarity 46.5%; Pred. No. 0.24;
Matches 132; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

```
Qy          4 ATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCA 63
              |||   ||| | || |  |||      | || | |  | ||  | | |||   ||| |
Db        354 ATATATAATTATATAATTAAATATAATTAAATATATAATTATATATAATTATATATAATT 295

Qy         64 AGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGT 123
              |    | ||  || | |  | |||  || |     |||||  || | | | |||  |  |
Db        294 AAATATATAATTATATATAATTAAACATATAATTATATATAATTATATATAATTAAACAT 235

Qy        124 TTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCT 183
              | ||||   | ||        | | | | | ||  ||  |     |||| | ||  |
Db        234 ATAATTAAATATAATTATATATAATTATATATAATTATATATAATTAATTAAATATAATT 175

Qy        184 TATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTG 243
              |||    | || | | || |  | |              |||    | |     |  |
Db        174 AATTATATATAATAAATATATAATTAAATATAATTAATTATATATAATAAATATATAATT 115

Qy        244 ACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGAT 287
              | || |   | |||  ||||  ||||  || |  ||| |  ||
Db        114 AAATATAATTAATTATATATAATAAATATATAATTAAATATAAT 71
```

RESULT 13
US-09-949-016-196366/c
; Sequence 196366, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 196366
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-196366

Query Match 11.1%; Score 40.8; DB 3; Length 601;
Best Local Similarity 46.5%; Pred. No. 0.24;
Matches 132; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

```
Qy          4 ATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCA 63
              |||   ||| | || |  |||      | || | |  | ||  | | |||   ||| |
Db        352 ATATATAATTATATAATTAAATATAATTAAATATATAATTATATATAATTATATATAATT 293

Qy         64 AGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGT 123
              |    | ||  || | |  | |||  || |     |||||  || | | | |||  |  |
Db        292 AAATATATAATTATATATAATTAAACATATAATTATATATAATTATATATAATTAAACAT 233
```

```
Qy        124 TTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCT 183
              | ||||   | |       | | | | | | ||  ||  |      |||| | ||  |
Db        232 ATAATTAAATATAATTATATATAATTATATATAATTATATATAATTAATTAAATATAATT 173

Qy        184 TATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTG 243
               |||   | || | | || |  | |        ||||   |  |     |    |
Db        172 AATTATATATAATAAATATATAATTAAATATAATTAATTATATATAATAAATATATAATT 113

Qy        244 ACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGAT 287
              | || |    | |||  ||||  |||| || |  ||| |  ||
Db        112 AAATATAATTAATTATATATAATAAATATATAATTAAATATAAT 69


RESULT 14
US-09-949-016-196367/c
; Sequence 196367, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 196367
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-196367

  Query Match             11.1%;  Score 40.8;  DB 3;  Length 601;
  Best Local Similarity   46.5%;  Pred. No. 0.24;
  Matches  132;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy          4 ATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCA 63
              |||   ||| | || |  |||     | || | |  | ||  | | |||   ||| |
Db        316 ATATATAATTATATAATTAAATATAATTAAATATATAATTATATATAATTATATATAATT 257

Qy         64 AGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGT 123
              |    | ||  || | |  | |||  || |    |||||  || | | | |||  | |
Db        256 AAATATATAATTATATATAATTAAACATATAATTATATATAATTATATATAATTAAACAT 197

Qy        124 TTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCT 183
              | ||||   | |       |  | | | | | ||  ||  |      |||| | ||  |
Db        196 ATAATTAAATATAATTATATATAATTATATATAATTATATATAATTAATTAAATATAATT 137

Qy        184 TATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTG 243
               |||    | || | | || |  | |         ||||    |  |     |    |
Db        136 AATTATATATAATAAATATATAATTAAATATAATTAATTATATATAATAAATATATAATT 77

Qy        244 ACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGAT 287
              | || |    | |||  ||||  ||||  || |  ||| |  ||
Db         76 AAATATAATTAATTATATATAATAAATATATAATTAAATATAAT 33


RESULT 15
US-09-949-016-12896/c
; Sequence 12896, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12896
;   LENGTH: 451924
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12896

  Query Match             11.1%;  Score 40.8;  DB 3;  Length 451924;
  Best Local Similarity   46.5%;  Pred. No. 1.3;
  Matches  132;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy          4 ATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCA 63
              |||   ||| | || |  |||     | || | |  | ||  | | |||   ||| |
Db     140767 ATATATAATTATATAATTAAATATAATTAAATATATAATTATATATAATTATATATAATT 140708

Qy         64 AGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGT 123
              |    | ||  || | |  | |||  || |    |||||  || | | | |||  | |
Db     140707 AAATATATAATTATATATAATTAAACATATAATTATATATAATTATATATAATTAAACAT 140648

Qy        124 TTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCT 183
              | ||||   | |       |  | | | | | ||  ||  |      |||| | ||  |
Db     140647 ATAATTAAATATAATTATATATAATTATATATAATTATATATAATTAATTAAATATAATT 140588

Qy        184 TATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTG 243
               |||    | || | | || |  | |         ||||    |  |     |    |
Db     140587 AATTATATATAATAAATATATAATTAAATATAATTAATTATATATAATAAATATATAATT 140528

Qy        244 ACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGAT 287
              | || |    | |||  ||||  ||||  || |  ||| |  ||
Db     140527 AAATATAATTAATTATATATAATAAATATATAATTAAATATAAT 140484


Search completed: May 11, 2006, 05:23:04
Job time : 54.0518 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:01:11 ; Search time 241.571 Seconds
(without alignments)
12528.801 Million cell updates/sec

Title: US-09-423-093B-2_COPY_2744_3109
Perfect score: 366
Sequence: 1 atgataatgaataaaatcaa.....tcactttgctggctggatta 366

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	366	100.0	14024	3	US-09-423-093A-2	Sequence 2, Appli
c	2	366	100.0	14187	5	US-10-114-170-121	Sequence 121, App
	3	47.2	12.9	647	4	US-09-925-065A-713654	Sequence 713654,
	4	46	12.6	647	4	US-09-925-065A-713652	Sequence 713652,
	5	46	12.6	647	4	US-09-925-065A-713653	Sequence 713653,
	6	46	12.6	647	4	US-09-925-065A-713655	Sequence 713655,
c	7	45.2	12.3	5542	6	US-10-311-455-1994	Sequence 1994, Ap
c	8	44.4	12.1	573	4	US-09-925-065A-398744	Sequence 398744,
c	9	44.4	12.1	573	4	US-09-925-065A-398745	Sequence 398745,
	10	43.8	12.0	625	4	US-09-925-065A-861747	Sequence 861747,
	11	43.8	12.0	646	4	US-09-925-065A-941413	Sequence 941413,
	12	43.8	12.0	6098	5	US-10-239-676-126	Sequence 126, App
	13	43.8	12.0	3673778	6	US-10-312-841-1	Sequence 1, Appli
	14	43.4	11.9	646	4	US-09-925-065A-941412	Sequence 941412,
	15	43.4	11.9	15768	6	US-10-311-455-1143	Sequence 1143, Ap
	16	42.8	11.7	9965	6	US-10-311-455-1500	Sequence 1500, Ap
	17	42.6	11.6	667	4	US-09-925-065A-553646	Sequence 553646,
c	18	42.6	11.6	41624	5	US-10-087-192-478	Sequence 478, App
c	19	42.4	11.6	6361	6	US-10-311-455-1113	Sequence 1113, Ap
	20	42.2	11.5	1856	4	US-09-925-065A-682004	Sequence 682004,
	21	42.2	11.5	6656	7	US-10-240-589C-118	Sequence 118, App
	22	42.2	11.5	8873	7	US-10-221-613-82	Sequence 82, Appl
c	23	42.2	11.5	3673778	6	US-10-312-841-2	Sequence 2, Appli
	24	42	11.5	2396	8	US-10-602-494-154	Sequence 154, App
	25	42	11.5	2396	8	US-10-602-494-296	Sequence 296, App
	26	42	11.5	40766	10	US-11-097-143-40270	Sequence 40270, A
	27	42	11.5	42429	10	US-11-097-143-24106	Sequence 24106, A
c	28	41.8	11.4	5748	6	US-10-311-455-1115	Sequence 1115, Ap
	29	41.6	11.4	11694	7	US-10-221-714A-421	Sequence 421, App
c	30	41.4	11.3	1021	3	US-09-764-853-34	Sequence 34, Appl
c	31	41.4	11.3	6071	6	US-10-311-455-298	Sequence 298, App
c	32	41.4	11.3	6071	7	US-10-221-613-32	Sequence 32, Appl
	33	41.2	11.3	1483	9	US-10-764-212-79	Sequence 79, Appl
c	34	41	11.2	526	8	US-10-425-115-109691	Sequence 109691,
	35	41	11.2	38342	7	US-10-221-714A-472	Sequence 472, App
c	36	40.8	11.1	535	8	US-10-357-930-55446	Sequence 55446, A
	37	40.8	11.1	6071	6	US-10-311-455-297	Sequence 297, App
	38	40.8	11.1	6071	7	US-10-221-613-31	Sequence 31, Appl
c	39	40.8	11.1	6222	6	US-10-311-455-666	Sequence 666, App
	40	40.6	11.1	6478	5	US-10-239-676-124	Sequence 124, App
	41	40.6	11.1	6478	6	US-10-240-453-144	Sequence 144, App
	42	40.6	11.1	6626	7	US-10-221-714A-536	Sequence 536, App
c	43	40.4	11.0	3673778	6	US-10-312-841-1	Sequence 1, Appli
	44	40.4	11.0	3673778	6	US-10-312-841-2	Sequence 2, Appli
	45	40.2	11.0	577	3	US-09-814-353-1167	Sequence 1167, Ap

ALIGNMENTS

RESULT 1
US-09-423-093A-2
; Sequence 2, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 14024
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-423-093A-2

Query Match 100.0%; Score 366; DB 3; Length 14024;
Best Local Similarity 100.0%; Pred. No. 4.2e-76;
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2744 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 2803

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2804 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 2863

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2864 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 2923

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2924 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 2983

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2984 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 3043
```

```
Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3044 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 3103

Qy        361 GGATTA 366
              ||||||
Db       3104 GGATTA 3109


RESULT 2
US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Publication No. US20030023075A1
;  GENERAL INFORMATION:
;       APPLICANT: Blattner, Frederick R.
;                  Burland,     Valerie
;                  Perna,  Nicole T.
;                  Plunkett, Guy
;                  Welch,  Rod
;       TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
;       NUMBER OF SEQUENCES: 265
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Quarles & Brady
;            STREET: 1 South Pinckney Street
;            CITY: Madison
;            STATE: WI
;            COUNTRY: US
;            ZIP: 53701-2113
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: Word Perfect 8.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/114,170
;            FILING DATE: 01-Apr-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/453,702
;            FILING DATE: 03-DEC-1999
;            APPLICATION NUMBER: 60/110,955
;            FILING DATE: 04-DEC-1998
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Seay, Nicholas J.
;            REGISTRATION NUMBER: 27386
;            REFERENCE/DOCKET NUMBER: 960296.95017
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (608) 251-5000
;            TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 121:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 14187
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA  (genomic)
;       SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121

  Query Match            100.0%;  Score 366;  DB 5;  Length 14187;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-76;
  Matches  366;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11257 ATGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACA 11198

Qy         61 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11197 TCAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCA 11138

Qy        121 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11137 AGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTA 11078

Qy        181 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11077 CCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCT 11018

Qy        241 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11017 CTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTT 10958

Qy        301 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10957 GCGTGTGGCAAAAATTTAAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCT 10898

Qy        361 GGATTA 366
              ||||||
Db      10897 GGATTA 10892


RESULT 3
US-09-925-065A-713654
; Sequence 713654, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 713654
;   LENGTH: 647
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-713654

  Query Match            12.9%;  Score 47.2;  DB 4;  Length 647;
  Best Local Similarity  49.6%;  Pred. No. 0.42;
  Matches  118;  Conservative    1;  Mismatches  119;  Indels    0;  Gaps    0;

Qy          2 TGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACAT 61
              | | ||  || |: |  ||||    ||      ||||      | || ||     |    | |
Db        380 TTAAAAAAAAAAMCAAGAAAACTCTAAAACTATTTGAAGATATACAATGCAAAACAAAAC 439

Qy         62 CAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAA 121
               ||        | | | |||||||  |     |  |||    |      ||| || |||||||
Db        440 AAAACAAACAATGAAAAGAACAGCAGCAATGAAGGCCCACGATTTTGGCTTTTATTTCCT 499

Qy        122 GTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTAC 181
              ||  |  ||   |||| |    || | ||   ||| | |||||  | | ||
Db        500 ATTCTACCAGGGGAATAAACATTTTATTCACCCTGGGATTTAGTGTAAAGTGAGCTAGTT 559

Qy        182 CTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGC 239
              |||||  ||||  ||||   ||| |       ||| | ||| |  |  ||||||
Db        560 AACATTTAACACAAGCCAGATGCTGTGGACCTTGTGCTAATTCCAATTCTTCTGCTGC 617


RESULT 4
US-09-925-065A-713652
; Sequence 713652, Application US/09925065A
```

```
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 713652
;   LENGTH: 647
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-713652

  Query Match             12.6%;  Score 46;  DB 4;  Length 647;
  Best Local Similarity   49.6%;  Pred. No. 0.81;
  Matches  118;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy          2 TGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACAT 61
              | | ||  || | |  |||   ||     ||||     | || ||    |    | |
Db        380 TTAAAAAAAAAACCAAGAAAACTCTAAAACTATTTGAAGATATACAATGCAAAACAAAAC 439

Qy         62 CAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAA 121
               ||       | | | ||||||| |    |  |||   |     ||| || |||||||
Db        440 AAAACAAACAATGAAAAGAACAGCAGCAATGAAGGCCCACGATTTTGGCTTTTATTTCCT 499

Qy        122 GTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTAC 181
               ||  |  ||   |||| |    || |  | ||   ||| | ||||| | | ||
Db        500 ATTCTACCAGGGGAATAAACATTTTATTCACCCTGGGATTTAGTGTAAAGTGAGCTAGTT 559

Qy        182 CTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGC 239
              |||||  |||||  ||||   ||| |       ||| | ||| |  |  ||||||
Db        560 AACATTTAACACAAGCCAGATGCTGTGGACCTTGTGCTAATTCCAATTCTTCTGCTGC 617


RESULT 5
US-09-925-065A-713653
; Sequence 713653, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 713653
;   LENGTH: 647
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-713653

  Query Match             12.6%;  Score 46;  DB 4;  Length 647;
  Best Local Similarity   49.6%;  Pred. No. 0.81;
  Matches  118;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy          2 TGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACAT 61
              | | ||  || | |  |||   ||     ||||     | || ||    |    | |
Db        380 TTAAAAAAAAAACCAAGAAAACTCTAAAACTATTTGAAGATATACAATGCAAAACAAAAC 439

Qy         62 CAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAA 121
               ||       | | | ||||||| |    |  |||   |     ||| || |||||||
Db        440 AAAACAAACAATGAAAAGAACAGCAGCAATGAAGGCCCACGATTTTGGCTTTTATTTCCT 499

Qy        122 GTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTAC 181
               ||  |  ||   |||| |    || |  | ||   ||| | ||||| | | ||
Db        500 ATTCTACCAGGGGAATAAACATTTTATTCACCCTGGGATTTAGTGTAAAGTGAGCTAGTT 559

Qy        182 CTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGC 239
              |||||  |||||  ||||   ||| |       ||| | ||| |  |  ||||||
Db        560 AACATTTAACACAAGCCAGATGCTGTGGACCTTGTGCTAATTCCAATTCTTCTGCTGC 617


RESULT 6
US-09-925-065A-713655
; Sequence 713655, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 713655
;   LENGTH: 647
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-713655

  Query Match             12.6%;  Score 46;  DB 4;  Length 647;
  Best Local Similarity   49.6%;  Pred. No. 0.81;
  Matches  118;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy          2 TGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACAT 61
              | | ||  || | |  |||   ||     ||||     | || ||    |    | |
Db        380 TTAAAAAAAAAACCAAGAAAACTCTAAAASTATTTGAAGATATACAATGCAAAACAAAAC 439

Qy         62 CAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAA 121
               ||       | | | ||||||| |    |  |||   |     ||| || |||||||
Db        440 AAAACAAACAATGAAAAGAACAGCAGCAATGAAGGCCCACGATTTTGGCTTTTATTTCCT 499

Qy        122 GTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTAC 181
               ||  |  ||   |||| |    || |  | ||   ||| | ||||| | | ||
Db        500 ATTCTACCAGGGGAATAAACATTTTATTCACCCTGGGATTTAGTGTAAAGTGAGCTAGTT 559

Qy        182 CTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGC 239
              |||||  |||||  ||||   ||| |       ||| | ||| |  |  ||||||
Db        560 AACATTTAACACAAGCCAGATGCTGTGGACCTTGTGCTAATTCCAATTCTTCTGCTGC 617
```

```
RESULT 7
US-10-311-455-1994/c
; Sequence 1994, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1994
;   LENGTH: 5542
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1994

  Query Match             12.3%;  Score 45.2;  DB 6;  Length 5542;
  Best Local Similarity   54.9%;  Pred. No. 2.9;
  Matches   89;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy          12 TAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCTTT 71
               |||| | ||||||| ||  |||||||  |   | || |||| || | |  |||| |   |
Db        3843 TAAATTAAAAAAAAAACAAAAATTTTACTAATCAATAAAAAATAAAATTTCAAATCTATA 3784

Qy          72 AGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAG 131
               |  ||   | |   |||   | |  |  | |   || |||   || | ||  |  |
Db        3783 AAATATTTACCTTAAAAAAAAAAAAAAAAAAAAAAATTATCTACCATATAAATCTAAAAATC 3724

Qy         132 TAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTT 173
               ||||  ||   |||||||   | || ||  ||  ||  ||
Db        3723 ATAAATTCTAAAACTACTTTAATTCATCTTTCCTAAAATATT 3682


RESULT 8
US-09-925-065A-398744/c
; Sequence 398744, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 398744
;   LENGTH: 573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-398744

  Query Match             12.1%;  Score 44.4;  DB 4;  Length 573;
  Best Local Similarity   51.5%;  Pred. No. 1.8;
  Matches  102;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy         109 TCTGTTATTTCAAGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTA 168
               | | ||  ||    || ||| | || |     |   | | ||   || || | ||  |  |
Db         382 TATTTTTATTTCTCTTAGATAAATACATAGTTGGAATTGCTGGATCATAGAGTAGTTTCA 323

Qy         169 AGTTTAACTTTACCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGT 228
               ||||||  ||||  |||   || |   ||  ||| |    ||    || |||  |  | |
Db         322 CGTTTAAATTTATATTAAAAAGAAACTGAACGATATTCCATAGTAGTGTCTAACAGAATT 263

Qy         229 CTTGGTGCTGCTCTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATC 288
               ||   ||    ||| | |||| || | ||       || |||  || | |||  ||   |
Db         262 CTCCTTGAGTTTCTCATATTTATGTATTTCTTACCAGCAAAAACACTGACAGCCATTGTT 203

Qy         289 GCACATTCATTTGCGTGT 306
               || | ||  ||  |||
Db         202 TCAAACTCTCTTATTTGT 185


RESULT 9
US-09-925-065A-398745/c
; Sequence 398745, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 398745
;   LENGTH: 573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-398745

  Query Match             12.1%;  Score 44.4;  DB 4;  Length 573;
  Best Local Similarity   51.5%;  Pred. No. 1.8;
  Matches  102;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy         109 TCTGTTATTTCAAGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTA 168
               | | ||  ||    || ||| | || |     |   | | ||   || || | ||  |  |
Db         382 TATTTTTATTTCTCTTAGATAAATACATAGTTGGAATTGCTGGATCATAGAGTAGTTTCA 323

Qy         169 AGTTTAACTTTACCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGT 228
               ||||||  ||||  |||   || |   ||  ||| |    ||    || |||  |  | |
Db         322 CGTTTAAATTTATATTAAAAAGAAACTGAACGATATTCCATAGTAGTGTCTAACAGAATT 263

Qy         229 CTTGGTGCTGCTCTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATC 288
               ||   ||    ||| | |||| || | ||       || |||  || | |||  ||   |
Db         262 CTCCTTGAGTTTCTCATATTTATGTATTTCTTACCAGCAAAAACACTGACAGCCATTGTT 203

Qy         289 GCACATTCATTTGCGTGT 306
               || | ||  ||  |||
Db         202 TCAAACTCTCTTATTTGT 185
```

```
RESULT 10
US-09-925-065A-861747
; Sequence 861747, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 861747
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-861747

  Query Match             12.0%;  Score 43.8;  DB 4;  Length 625;
  Best Local Similarity   47.9%;  Pred. No. 2.7;
  Matches  123;  Conservative    1;  Mismatches  133;  Indels    0;  Gaps    0;

Qy         10 AATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCT 69
              || ||| || || | ||   |||||||| |||| || ||| | | | :| |||     |
Db        289 AACAAATTCTAACAGATTTATAAATTTTAAACTTCAACAAATTTTAAAWCCTCATTATTT 348

Qy         70 TTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATT 129
              |||  | |   ||||   | |       | | |     |      || | |    | |||||
Db        349 TTATTTTGCTTACAGCTCATCTCACTTCTGAGAAAAATTAACTTTTTATACTATCTGATT 408

Qy        130 AGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTA 189
              || ||| ||| |   |  | ||| |   |  |||| |   |     ||  |
Db        409 TCAAATATAATCTGATCTAATAATTTTAATTGTTCCTTAAGCATTGTTCACTTTTTCTGG 468

Qy        190 GGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACATTT 249
              |   || | |       | || |     |||||  |   || | || | || |  |||
Db        469 GATAAAAAAATTCAGTATTTAAGATCATCTATTTTCTTCCTAGTTGTTTCTGTACCATGA 528

Qy        250 TTGGACTTAGGTATAGG 266
              |   |  |  ||||| |
Db        529 TGACATCTGAGTATATG 545


RESULT 11
US-09-925-065A-941413
; Sequence 941413, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 941413
;   LENGTH: 646
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-941413

  Query Match             12.0%;  Score 43.8;  DB 4;  Length 646;
  Best Local Similarity   51.8%;  Pred. No. 2.7;
  Matches   99;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          8 TGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTG 67
              | ||||| ||| | | |   |  | | ||  |   | | | | ||| || |||    |
Db        230 TTAATAAGATCTATATTAATATATATTATTATATAATTATATACTATTATTCATATTATT 289

Qy         68 CTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGA 127
                ||  || |  | |  ||| || |  ||  |||  |||| |  ||||   | | | |
Db        290 AAATATATATAATATATTAAATATATAATATAGTATAATTGTATAATATTAACAATCTAA 349

Qy        128 TTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATT 187
              ||| ||  |||   | |       | | | ||||  ||  | | | ||  |||    |||
Db        350 TTAATATTATATAATTATACTAATATATAATATATACATTTAATATATATTTGATATATA 409

Qy        188 TAGGACAAGAG 198
              || |  ||| |
Db        410 TAYGTAAAGGG 420


RESULT 12
US-10-239-676-126
; Sequence 126, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 126
;   LENGTH: 6098
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-126

  Query Match             12.0%;  Score 43.8;  DB 5;  Length 6098;
  Best Local Similarity   54.7%;  Pred. No. 6.5;
  Matches   87;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         72 AGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAG 131
              |||||||| | | ||| ||||   | |||| |   |  |  ||||  | || ||   |  |
Db       2103 AGGTAGAGGAGAAGAATGGTAGTTGTTTATTTTGTGTTAGGTTAGGTTAATTTATTTATG 2162

Qy        132 TAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTAGG 191
              || | || |      |  ||  | | |||||||      || |    ||    ||||||  |
Db       2163 TATATTATTTGATTTGGTTATTATAATATTAAGAAATAGATATTATTATTTTTATTTTAG 2222
```

```
Qy        192 ACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCT 230
              |   |||  | ||||| ||  ||    | ||   ||| |
Db       2223 AGTCGAGGAAATTGGTTTAGAGAGATTTGTTCGAAGTTT 2261
```

RESULT 13
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 12.0%; Score 43.8; DB 6; Length 3673778;
Best Local Similarity 51.3%; Pred. No. 79;
Matches 102; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```
Qy         85 GAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAGTAAAATACTCTCA 144
              ||||  |  | ||  | |       | || ||||| || |  || || |||       |
Db    1597809 GAAATATGAAAGAGAAAAAGTAGTTATGATATTTTAAATAAGAATAATAAGTAGAGAGAA 1597868

Qy        145 CTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTAGGACAAGAGAGATTT 204
              |  |||       |||||||   | |  |||   |||  ||| ||| || |   |||||
Db    1597869 TTTTTTTAGTAGATATTAAGATTTATGTTAGAGTTATTTTAATTAAGAGAGTCGAGATGT 1597928

Qy        205 GGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACATTTTTGGACTTAGGTATA 264
              | |   || |    | |   | | | ||     |    || | ||||| || || ||| |
Db    1597929 TGAGATAGGGTAGTAAATTAGAATTTGGGAATAGAATAGAAAGTTTTGAACGTATGTAGA 1597988

Qy        265 GGAAATGCATTAACAAACA 283
              | | | | | ||  | | |
Db    1597989 GTAGAGGTAATATTATAGA 1598007
```

RESULT 14
US-09-925-065A-941412
; Sequence 941412, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 941412
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-941412

Query Match 11.9%; Score 43.4; DB 4; Length 646;
Best Local Similarity 51.3%; Pred. No. 3.3;
Matches 98; Conservative 1; Mismatches 92; Indels 0; Gaps 0;

```
Qy          8 TGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTG 67
              | ||||| ||| | | |   |  | | ||  |   | | | | ||| || |||     |
Db        230 TTAATAAGATCTATATTAATATATTATTATATAATTATATACTATTATTCATATTATT 289

Qy         68 CTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGA 127
                ||  || |  | |  |||  || |  ||  |||  |||| |  ||||   | | | |
Db        290 AAATATATATAATATATTAAATATATAATATAGTATAATTGTATAATATTAACAATCTAA 349

Qy        128 TTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATT 187
              ||: ||  |||   | |       | | | ||||  ||  | | | ||  |||    |||
Db        350 TTRATATTATATAATTATACTAATATATAATATATACATTTAATATATATTTGATATATA 409

Qy        188 TAGGACAAGAG 198
              || |  ||| |
Db        410 TATGTAAAGGG 420
```

RESULT 15
US-10-311-455-1143
; Sequence 1143, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1143
; LENGTH: 15768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1143

Query Match 11.9%; Score 43.4; DB 6; Length 15768;
Best Local Similarity 48.9%; Pred. No. 12;
Matches 116; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

```
Qy         29 TTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCTTTAGGTAGAGAACAGGAAA 88
              ||  | ||     ||| |    || || |   |  ||  ||||      |  | | |
Db      15340 TTCGAATTGTTGATTTTAGGTGATTTGTTTATTTCGGTTTTTTAAAGTGTTGGGATTATA 15399

Qy         89 GGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATTAGTAAAATACTCTCACTAC 148
              ||||   | |  |||  ||   |||  | |    ||| ||  || ||||| | |   |
Db      15400 GGTATGAGTTATTATGTTTAGTTGTAGTATAGGTTTTAATATGTTAAATATTTTTTTTTT 15459

Qy        149 TTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTAGGACAAGAGAGATTTGGTG 208
              |||    | || |  |  | |  |||||  ||   || ||        ||||||  ||   |
Db      15460 TTTTATTAATGTGTATGGAAGTTTTAATATTTTTTTTTTATATTTTAGAGAGTTTATATT 15519

Qy        209 TATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACATTTTTGGACTTAGGTATAG 265
              |  | || | || ||  ||| || ||        ||   |||||||  ||   | | |
Db      15520 TTGGAATTAATAATATTAGTTTTTGTTGATAAGTGTTTTTTTTGGGTTTTTTTTTTG 15576
```

Search completed: May 11, 2006, 07:08:54
Job time : 251.571 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:23:58 ; Search time 189.131 Seconds
(without alignments)
7892.621 Million cell updates/sec

Title: US-09-423-093B-2_COPY_2744_3109
Perfect score: 366
Sequence: 1 atgataatgaataaaatcaa.....tcactttgctggctggatta 366

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	47.2	12.9	647	7	US-09-925-065A-713654	Sequence 713654,
	2	46	12.6	647	7	US-09-925-065A-713652	Sequence 713652,
	3	46	12.6	647	7	US-09-925-065A-713653	Sequence 713653,
	4	46	12.6	647	7	US-09-925-065A-713655	Sequence 713655,
c	5	44.4	12.1	573	7	US-09-925-065A-398744	Sequence 398744,
c	6	44.4	12.1	573	7	US-09-925-065A-398745	Sequence 398745,
c	7	44.4	12.1	593	12	US-10-301-480-466143	Sequence 466143,
c	8	44.4	12.1	593	12	US-10-301-480-466144	Sequence 466144,
c	9	44.4	12.1	593	12	US-10-301-480-1079552	Sequence 1079552,
c	10	44.4	12.1	593	12	US-10-301-480-1079553	Sequence 1079553,
	11	43.8	12.0	625	7	US-09-925-065A-861747	Sequence 861747,
	12	43.8	12.0	646	7	US-09-925-065A-941413	Sequence 941413,
c	13	43.8	12.0	881	12	US-10-301-480-541537	Sequence 541537,
c	14	43.8	12.0	881	12	US-10-301-480-1154946	Sequence 1154946,
	15	43.4	11.9	646	7	US-09-925-065A-941412	Sequence 941412,
	16	42.6	11.6	667	7	US-09-925-065A-553646	Sequence 553646,
	17	42.2	11.5	1856	7	US-09-925-065A-682004	Sequence 682004,
	18	42.2	11.5	6656	10	US-10-240-708-76	Sequence 76, Appl
	19	40.8	11.1	1739	10	US-10-750-185-63166	Sequence 63166, A
	20	40.8	11.1	1739	10	US-10-750-623-63166	Sequence 63166, A
	21	40.2	11.0	179597	17	US-11-121-086-91	Sequence 91, Appl
	22	39.4	10.8	721	12	US-10-301-480-585469	Sequence 585469,
	23	39.4	10.8	721	12	US-10-301-480-1198878	Sequence 1198878,
	24	39.2	10.7	654	7	US-09-925-065A-867324	Sequence 867324,
c	25	39	10.7	636	7	US-09-925-065A-858185	Sequence 858185,
c	26	39	10.7	1495	7	US-09-925-065A-94375	Sequence 94375, A
c	27	39	10.7	1495	7	US-09-925-065A-94376	Sequence 94376, A
c	28	39	10.7	1495	11	US-10-301-480-195617	Sequence 195617,
c	29	39	10.7	1495	11	US-10-301-480-195618	Sequence 195618,
c	30	39	10.7	1495	12	US-10-301-480-809026	Sequence 809026,
c	31	39	10.7	1495	12	US-10-301-480-809027	Sequence 809027,
c	32	39	10.7	1557	7	US-09-925-065A-555017	Sequence 555017,
c	33	38.8	10.6	542	11	US-10-301-480-12218	Sequence 12218, A
c	34	38.8	10.6	542	12	US-10-301-480-625627	Sequence 625627,
	35	38.8	10.6	2222	10	US-10-750-185-37563	Sequence 37563, A
	36	38.8	10.6	2222	10	US-10-750-623-37563	Sequence 37563, A
c	37	38.8	10.6	2944528	18	US-11-045-004-1	Sequence 1, Appli
	38	38.6	10.5	486	7	US-09-925-065A-12278	Sequence 12278, A
	39	38.6	10.5	486	11	US-10-301-480-113515	Sequence 113515,
	40	38.6	10.5	486	12	US-10-301-480-726924	Sequence 726924,
c	41	38.6	10.5	544	7	US-09-925-065A-31218	Sequence 31218, A
c	42	38.6	10.5	544	11	US-10-301-480-132456	Sequence 132456,
c	43	38.6	10.5	544	12	US-10-301-480-745865	Sequence 745865,
	44	38.6	10.5	552	7	US-09-925-065A-399559	Sequence 399559,
	45	38.6	10.5	561	12	US-10-301-480-466917	Sequence 466917,

ALIGNMENTS

RESULT 1
US-09-925-065A-713654
; Sequence 713654, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 713654
; LENGTH: 647
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-713654

Query Match 12.9%; Score 47.2; DB 7; Length 647;
Best Local Similarity 49.6%; Pred. No. 2;
Matches 118; Conservative 1; Mismatches 119; Indels 0; Gaps 0;

Qy 2 TGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACAT 61
Db 380 TTAAAAAAAAAAMCAAGAAAACTCTAAAACTATTTGAAGATATACAATGCAAAACAAAAC 439

```
Qy          62 CAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAA 121
               ||      | | | |||||||| |    |  |||   |     ||| || |||||||
Db         440 AAAACAAACAATGAAAAGAACAGCAGCAATGAAGGCCCACGATTTTGGCTTTTATTTCCT 499

Qy         122 GTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTAC 181
               ||  |  ||   ||||  |    || |  | ||   ||| | |||||  | | ||
Db         500 ATTCTACCAGGGGAATAAACATTTTATTCACCCTGGGATTTAGTGTAAAGTGAGCTAGTT 559

Qy         182 CTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGC 239
               |||||  |||||  ||||   ||| |       ||| | ||| |  |  ||||||
Db         560 AACATTTAACACAAGCCAGATGCTGTGGACCTTGTGCTAATTCCAATTCTTCTGCTGC 617


RESULT 2
US-09-925-065A-713652
; Sequence 713652, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 713652
;   LENGTH: 647
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-713652

  Query Match             12.6%;  Score 46;  DB 7;  Length 647;
  Best Local Similarity   49.6%;  Pred. No. 3.6;
  Matches  118;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy           2 TGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACAT 61
               | | ||  || |  | ||||   ||     ||||     | ||||     |   | |
Db         380 TTAAAAAAAAAACCAAGAAAACTCTAAAACTATTTGAAGATATACAATGCAAAACAAAAC 439

Qy          62 CAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAA 121
               ||      | | |||||||| |    |  |||   |     ||| || |||||||
Db         440 AAAACAAACAATGAAAAGAACAGCAGCAATGAAGGCCCACGATTTTGGCTTTTATTTCCT 499

Qy         122 GTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTAC 181
               ||  |  ||   ||||  |    || |  | ||   ||| | |||||  | | ||
Db         500 ATTCTACCAGGGGAATAAACATTTTATTCACCCTGGGATTTAGTGTAAAGTGAGCTAGTT 559

Qy         182 CTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGC 239
               |||||  |||||  ||||   ||| |       ||| | ||| |  |  ||||||
Db         560 AACATTTAACACAAGCCAGATGCTGTGGACCTTGTGCTAATTCCAATTCTTCTGCTGC 617


RESULT 3
US-09-925-065A-713653
; Sequence 713653, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 713653
;   LENGTH: 647
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-713653

  Query Match             12.6%;  Score 46;  DB 7;  Length 647;
  Best Local Similarity   49.6%;  Pred. No. 3.6;
  Matches  118;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy           2 TGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACAT 61
               | | ||  || |  | ||||   ||     ||||     | ||||     |   | |
Db         380 TTAAAAAAAAAACCAAGAAAACTCTAAAACTATTTGAAGATATACAATGCAAAACAAAAC 439

Qy          62 CAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAA 121
               ||      | | |||||||| |    |  |||   |     ||| || |||||||
Db         440 AAAACAAACAATGAAAAGAACAGCAGCAATGAAGGCCCACGATTTTGGCTTTTATTTCCT 499

Qy         122 GTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTAC 181
               ||  |  ||   ||||  |    || |  | ||   ||| | |||||  | | ||
Db         500 ATTCTACCAGGGGAATAAACATTTTATTCACCCTGGGATTTAGTGTAAAGTGAGCTAGTT 559

Qy         182 CTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGC 239
               |||||  |||||  ||||   ||| |       ||| | ||| |  |  ||||||
Db         560 AACATTTAACACAAGCCAGATGCTGTGGACCTTGTGCTAATTCCAATTCTTCTGCTGC 617


RESULT 4
US-09-925-065A-713655
; Sequence 713655, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 713655
;   LENGTH: 647
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-713655

  Query Match             12.6%;  Score 46;  DB 7;  Length 647;
  Best Local Similarity   49.6%;  Pred. No. 3.6;
  Matches  118;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;
```

```
Qy          2 TGATAATGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACAT 61
              | | ||  || |  |  ||||   ||     ||||     | || ||    |    | |
Db        380 TTAAAAAAAAAACCAAGAAAACTCTAAAASTATTTGAAGATATACAATGCAAAACAAAAC 439

Qy         62 CAAGTGCTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAA 121
               ||       | | | ||||||| |    |  |||   |      ||| || |||||||
Db        440 AAAACAAACAATGAAAAGAACAGCAGCAATGAAGGCCCACGATTTTGGCTTTTATTTCCT 499

Qy        122 GTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTAC 181
               ||  |  ||   ||||  |   || |  | ||   ||| | |||||  | | ||
Db        500 ATTCTACCAGGGGAATAAACATTTTATTCACCCTGGGATTTAGTGTAAAGTGAGCTAGTT 559

Qy        182 CTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGC 239
              |||||  |||||  ||||   ||| |       ||| | ||| |  |  ||||||
Db        560 AACATTTAACACAAGCCAGATGCTGTGGACCTTGTGCTAATTCCAATTCTTCTGCTGC 617


RESULT 5
US-09-925-065A-398744/c
; Sequence 398744, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 398744
;   LENGTH: 573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-398744

  Query Match             12.1%;  Score 44.4;  DB 7;  Length 573;
  Best Local Similarity   51.5%;  Pred. No. 7.4;
  Matches  102;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        109 TCTGTTATTTCAAGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTA 168
              | | ||  ||    || ||| | || |    |   | | ||   || || | ||  |  |
Db        382 TATTTTTATTTCTCTTAGATAAATACATAGTTGGAATTGCTGGATCATAGAGTAGTTTCA 323

Qy        169 AGTTTAACTTTACCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGT 228
              ||||||  ||||  |||   || |   ||  ||| |    ||    || |||  |  | |
Db        322 CGTTTAAATTTATATTAAAAAGAAACTGAACGATATTCCATAGTAGTGTCTAACAGAATT 263

Qy        229 CTTGGTGCTGCTCTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATC 288
              ||   ||    ||| | |||| || | ||      || |||  || | |||  ||   |
Db        262 CTCCTTGAGTTTCTCATATTTATGTATTTCTTACCAGCAAAAACACTGACAGCCATTGTT 203

Qy        289 GCACATTCATTTGCGTGT 306
              || | ||  ||   |||
Db        202 TCAAACTCTCTTATTTGT 185


RESULT 6
US-09-925-065A-398745/c
; Sequence 398745, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 398745
;   LENGTH: 573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-398745

  Query Match             12.1%;  Score 44.4;  DB 7;  Length 573;
  Best Local Similarity   51.5%;  Pred. No. 7.4;
  Matches  102;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        109 TCTGTTATTTCAAGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTA 168
              | | ||  ||    || ||| | || |    |   | | ||   || || | ||  |  |
Db        382 TATTTTTATTTCTCTTAGATAAATACATAGTTGGAATTGCTGGATCATAGAGTAGTTTCA 323

Qy        169 AGTTTAACTTTACCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGT 228
              ||||||  ||||  |||   || |   ||  ||| |    ||    || |||  |  | |
Db        322 CGTTTAAATTTATATTAAAAAGAAACTGAACGATATTCCATAGTAGTGTCTAACAGAATT 263

Qy        229 CTTGGTGCTGCTCTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATC 288
              ||   ||    ||| | |||| || | ||      || |||  || | |||  ||   |
Db        262 CTCCTTGAGTTTCTCATATTTATGTATTTCTTACCAGCAAAAACACTGACAGCCATTGTT 203

Qy        289 GCACATTCATTTGCGTGT 306
              || | ||  ||   |||
Db        202 TCAAACTCTCTTATTTGT 185


RESULT 7
US-10-301-480-466143/c
; Sequence 466143, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 466143
;   LENGTH: 593
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-466143

  Query Match             12.1%;  Score 44.4;  DB 12;  Length 593;
  Best Local Similarity   51.5%;  Pred. No. 7.5;
  Matches  102;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        109 TCTGTTATTTCAAGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTA 168
              | | ||  ||    || ||| | || |    |   | | ||   || || | ||  |  |
```

```
Db        403 TATTTTTATTTCTCTTAGATAAATACATAGTTGGAATTGCTGGATCATAGAGTAGTTTCA 344

Qy        169 AGTTTAACTTTACCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGT 228
              |||||| ||||  |||   || |   ||  ||| |     ||     || ||| |  | |
Db        343 CGTTTAAATTTATATTAAAAAGAAACTGAACGATATTCCATAGTAGTGTCTAACAGAATT 284

Qy        229 CTTGGTGCTGCTCTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATC 288
              ||   ||     ||| | |||| || | ||        || |||  || | |||  ||    |
Db        283 CTCCTTGAGTTTCTCATATTTATGTATTTCTTACCAGCAAAAACACTGACAGCCATTGTT 224

Qy        289 GCACATTCATTTGCGTGT 306
              || | ||  ||   |||
Db        223 TCAAACTCTCTTATTTGT 206


RESULT 8
US-10-301-480-466144/c
; Sequence 466144, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 466144
;   LENGTH: 593
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-466144

  Query Match             12.1%;  Score 44.4;  DB 12;  Length 593;
  Best Local Similarity   51.5%;  Pred. No. 7.5;
  Matches  102;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        109 TCTGTTATTTCAAGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTA 168
              | | ||  ||    || ||| | || |    |  | | ||   || || | ||  | |
Db        403 TATTTTTATTTCTCTTAGATAAATACATAGTTGGAATTGCTGGATCATAGAGTAGTTTCA 344

Qy        169 AGTTTAACTTTACCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGT 228
              |||||| ||||  |||   || |   ||  ||| |     ||     || ||| |  | |
Db        343 CGTTTAAATTTATATTAAAAAGAAACTGAACGATATTCCATAGTAGTGTCTAACAGAATT 284

Qy        229 CTTGGTGCTGCTCTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATC 288
              ||   ||     ||| | |||| || | ||        || |||  || | |||  ||    |
Db        283 CTCCTTGAGTTTCTCATATTTATGTATTTCTTACCAGCAAAAACACTGACAGCCATTGTT 224

Qy        289 GCACATTCATTTGCGTGT 306
              || | ||  ||   |||
Db        223 TCAAACTCTCTTATTTGT 206


RESULT 9
US-10-301-480-1079552/c
; Sequence 1079552, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1079552
;   LENGTH: 593
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1079552

  Query Match             12.1%;  Score 44.4;  DB 12;  Length 593;
  Best Local Similarity   51.5%;  Pred. No. 7.5;
  Matches  102;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        109 TCTGTTATTTCAAGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTA 168
              | | ||  ||    || ||| | || |    |  | | ||   || || | ||  | |
Db        403 TATTTTTATTTCTCTTAGATAAATACATAGTTGGAATTGCTGGATCATAGAGTAGTTTCA 344

Qy        169 AGTTTAACTTTACCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGT 228
              |||||| ||||  |||   || |   ||  ||| |     ||     || ||| |  | |
Db        343 CGTTTAAATTTATATTAAAAAGAAACTGAACGATATTCCATAGTAGTGTCTAACAGAATT 284

Qy        229 CTTGGTGCTGCTCTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATC 288
              ||   ||     ||| | |||| || | ||        || |||  || | |||  ||    |
Db        283 CTCCTTGAGTTTCTCATATTTATGTATTTCTTACCAGCAAAAACACTGACAGCCATTGTT 224

Qy        289 GCACATTCATTTGCGTGT 306
              || | ||  ||   |||
Db        223 TCAAACTCTCTTATTTGT 206


RESULT 10
US-10-301-480-1079553/c
; Sequence 1079553, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1079553
;   LENGTH: 593
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1079553

  Query Match             12.1%;  Score 44.4;  DB 12;  Length 593;
  Best Local Similarity   51.5%;  Pred. No. 7.5;
  Matches  102;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        109 TCTGTTATTTCAAGTTTGATTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTA 168
              | | ||  ||    || ||| | || |    |  | | ||   || || | ||  | |
Db        403 TATTTTTATTTCTCTTAGATAAATACATAGTTGGAATTGCTGGATCATAGAGTAGTTTCA 344

Qy        169 AGTTTAACTTTACCTTATTTAGGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGT 228
              |||||| ||||  |||   || |   ||  ||| |     ||     || ||| |  | |
Db        343 CGTTTAAATTTATATTAAAAAGAAACTGAACGATATTCCATAGTAGTGTCTAACAGAATT 284

Qy        229 CTTGGTGCTGCTCTGACATTTTTGGACTTAGGTATAGGAAATGCATTAACAAACAGGATC 288
              ||   ||     ||| | |||| || | ||        || |||  || | |||  ||    |
Db        283 CTCCTTGAGTTTCTCATATTTATGTATTTCTTACCAGCAAAAACACTGACAGCCATTGTT 224

Qy        289 GCACATTCATTTGCGTGT 306
              || | ||  ||   |||
Db        223 TCAAACTCTCTTATTTGT 206
```

```
RESULT 11
US-09-925-065A-861747
; Sequence 861747, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 861747
;   LENGTH: 625
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-861747

  Query Match             12.0%;  Score 43.8;  DB 7;  Length 625;
  Best Local Similarity   47.9%;  Pred. No. 10;
  Matches  123;  Conservative    1;  Mismatches  133;  Indels    0;  Gaps    0;

Qy          10 AATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTGCT 69
               || ||| || || | ||   |||||||| |||| || ||| | | | :| |||     |
Db         289 AACAAATTCTAACAGATTTATAAATTTTAAACTTCAACAAATTTTAAAWCCTCATTATTT 348

Qy          70 TTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGATT 129
               |||  | |   ||||  | |      | | |     |      || | |    | |||||
Db         349 TTATTTTGCTTACAGCTCATCTCACTTCTGAGAAAAATTAACTTTTTATACTATCTGATT 408

Qy         130 AGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATTTA 189
                 || ||| ||| |     |  | |||  |   |  ||||  |   |     || |
Db         409 TCAAATATAATCTGATCTAATAATTTTAATTGTTCCTTAAGCATTGTTCACTTTTTCTGG 468

Qy         190 GGACAAGAGAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACATTT 249
               |   || | |       | || |     |||||  |   || | || | || |  |||
Db         469 GATAAAAAAAATTCAGTATTTAAGATCATCTATTTTCTTCCTAGTTGTTTCTGTACCATGA 528

Qy         250 TTGGACTTAGGTATAGG 266
               |   |  |  ||||| |
Db         529 TGACATCTGAGTATATG 545


RESULT 12
US-09-925-065A-941413
; Sequence 941413, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 941413
;   LENGTH: 646
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-941413

  Query Match             12.0%;  Score 43.8;  DB 7;  Length 646;
  Best Local Similarity   51.8%;  Pred. No. 10;
  Matches   99;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy           8 TGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTG 67
               | ||||| ||| | |   |    |  | | ||  |   | | | | ||| || |||    |
Db         230 TTAATAAGATCTATATTAATATATATTATTATATAATTATATACTATTATTCATATTATT 289

Qy          68 CTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGA 127
                 ||  || |  | |  |||  || |  ||  |||  |||| |  ||||   | | | |
Db         290 AAATATATATAATATATTAAATATATAATATAGTATAATTGTATAATATTAACAATCTAA 349

Qy         128 TTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATT 187
               ||| ||  |||   | |       | | | ||||  ||  | | | ||  |||    |||
Db         350 TTAATATTATATAATTATACTAATATATAATATATACATTTAATATATATTTGATATATA 409

Qy         188 TAGGACAAGAG 198
               || |  ||| |
Db         410 TAYGTAAAGGG 420


RESULT 13
US-10-301-480-541537/c
; Sequence 541537, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 541537
;   LENGTH: 881
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-541537

  Query Match             12.0%;  Score 43.8;  DB 12;  Length 881;
  Best Local Similarity   51.8%;  Pred. No. 11;
  Matches   99;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy           8 TGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTG 67
               | ||||| ||| | |   |    |  | | ||  |   | | | | ||| || |||    |
Db         417 TTAATAAGATCTATATTAATATATATTATTATATAATTATATACTATTATTCATATTATT 358

Qy          68 CTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGA 127
                 ||  || |  | |  |||  || |  ||  |||  |||| |  ||||   | | | |
Db         357 AAATATATATAATATATTAAATATATAATATAGTATAATTGTATAATATTAACAATCTAA 298

Qy         128 TTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATT 187
               ||| ||  |||   | |       | | | ||||  ||  | | | ||  |||    |||
Db         297 TTAATATTATATAATTATACTAATATATAATATATACATTTAATATATATTTGATATATA 238

Qy         188 TAGGACAAGAG 198
               || |  ||| |
```

```
Db        237 TATGTAAAGGG 227


RESULT 14
US-10-301-480-1154946/c
; Sequence 1154946, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1154946
;   LENGTH: 881
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1154946

  Query Match             12.0%;  Score 43.8;  DB 12;  Length 881;
  Best Local Similarity   51.8%;  Pred. No. 11;
  Matches   99;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          8 TGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTG 67
              | ||||| ||| | |  |   |  | | ||  |   | | | | ||| || |||    |
Db        417 TTAATAAGATCTATATTAATATATATTATTATATAATTATATACTATTATTCATATTATT 358

Qy         68 CTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGA 127
                ||  || |  | |  |||  || |  ||  |||  |||| |  ||||   | | | |
Db        357 AAATATATATAATATATTAAATATATAATATAGTATAATTGTATAATATTAACAATCTAA 298

Qy        128 TTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATT 187
              ||| ||  |||   | |      | | | ||||  ||  | | | ||  |||    |||
Db        297 TTAATATTATATAATTATACTAATATATAATATATACATTTAATATATATTTGATATATA 238

Qy        188 TAGGACAAGAG 198
              || |  ||| |
Db        237 TATGTAAAGGG 227


RESULT 15
US-09-925-065A-941412
; Sequence 941412, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 941412
;   LENGTH: 646
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-941412

  Query Match             11.9%;  Score 43.4;  DB 7;  Length 646;
  Best Local Similarity   51.3%;  Pred. No. 12;
  Matches   98;  Conservative    1;  Mismatches   92;  Indels    0;  Gaps    0;

Qy          8 TGAATAAAATCAAAAAAATACTTAAATTTTGCACTTTAAAAAAATATGATACATCAAGTG 67
              | ||||| ||| | |  |   |  | | ||  |   | | | | ||| || |||    |
Db        230 TTAATAAGATCTATATTAATATATATTATTATATAATTATATACTATTATTCATATTATT 289

Qy         68 CTTTAGGTAGAGAACAGGAAAGGTACAGGATTATATCCTTGTCTGTTATTTCAAGTTTGA 127
                ||  || |  | |  |||  || |  ||  |||  |||| |  ||||   | | | |
Db        290 AAATATATATAATATATTAAATATATAATATAGTATAATTGTATAATATTAACAATCTAA 349

Qy        128 TTAGTAAAATACTCTCACTACTTTCTCTTATATTAACTGTAAGTTTAACTTTACCTTATT 187
              ||: ||  |||   | |      | | | ||||  ||  | | | ||  |||    |||
Db        350 TTRATATTATATAATTATACTAATATATAATATATACATTTAATATATATTTGATATATA 409

Qy        188 TAGGACAAGAG 198
              || |  ||| |
Db        410 TATGTAAAGGG 420


Search completed: May 11, 2006, 08:30:27
Job time : 191.131 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:45:22 ; Search time 4301.31 Seconds
(without alignments)
15660.244 Million cell updates/sec

Title: US-09-423-093B-2_COPY_858_2042
Perfect score: 1185
Sequence: 1 gtgaagtcagcggctaagtt.....gttataactgtgacttataa 1185

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1185	100.0	14002	1	AF061251	AF061251 Escherici
2	1185	100.0	14024	6	BD085555	BD085555 Nucleic a
3	1185	100.0	14029	6	AX029262	AX029262 Sequence
c 4	1183.4	99.9	14181	6	BD184789	BD184789 Nucleic a
c 5	1183.4	99.9	14187	6	AR204225	AR204225 Sequence
c 6	1183.4	99.9	14187	6	AR637574	AR637574 Sequence
7	1183.4	99.9	31482	1	AB008676	AB008676 Escherich
c 8	1183.4	99.9	110000	1	AE005174_28	Continuation (29 o
c 9	1183.4	99.9	110000	1	BA000007_27	Continuation (28 o
10	521.6	44.0	14323	1	AY730593	AY730593 Salmonell
11	451.2	38.1	14224	1	AY730592	AY730592 Citrobact
12	82.8	7.0	190356	14	CR847965	CR847965 Danio rer
13	80.6	6.8	165165	14	CR628390	CR628390 Danio rer
c 14	80	6.8	175370	8	AC105933	AC105933 Homo sapi
15	77.4	6.5	1141	6	AR579680	AR579680 Sequence
16	77.4	6.5	1141	6	AX083744	AX083744 Sequence
c 17	77.4	6.5	110000	2	PFMAL1P2_1	Continuation (2 of
c 18	77.4	6.5	110000	2	PFMAL1P2_2	Continuation (3 of
19	76.4	6.4	110000	14	CT009752_3	Continuation (4 of
c 20	76.2	6.4	110000	14	PFMAL8P1_06	Continuation (7 of
21	75.8	6.4	4529	2	DQ054838	DQ054838 Plasmodiu
22	75.6	6.4	231489	14	AC156865	AC156865 Bos tauru
23	75.2	6.3	115218	8	HS159A1	AL034397 Human DNA
c 24	74.8	6.3	115218	8	HS159A1	AL034397 Human DNA
c 25	74.6	6.3	135121	8	AC069525	AC069525 Homo sapi
26	74.4	6.3	104992	14	AC005504	AC005504 Plasmodiu
27	74.4	6.3	169546	14	AC004157	AC004157 Plasmodiu
c 28	74.4	6.3	250421	2	AE014849	AE014849 Plasmodiu
c 29	74.2	6.3	180861	14	CR391970	CR391970 Danio rer
c 30	74	6.2	330050	2	PFA929355	AL929355 Plasmodiu
c 31	73.6	6.2	232202	14	CR769785	CR769785 Danio rer
c 32	73.6	6.2	250029	2	AE014839	AE014839 Plasmodiu
c 33	73.2	6.2	14867	2	AE001398	AE001398 Plasmodiu
c 34	73.2	6.2	176147	5	AL929105	AL929105 Zebrafish
35	73.2	6.2	226847	14	BX664721	BX664721 Danio rer
36	73	6.2	170627	14	AC125567	AC125567 Rattus no
c 37	72.8	6.1	333321	2	AC116986	AC116986 Dictyoste
c 38	72.6	6.1	108767	5	CR735141	CR735141 Zebrafish
c 39	72.6	6.1	165165	14	CR628390	CR628390 Danio rer
40	72.4	6.1	135011	14	CR381579	CR381579 Danio rer
c 41	72.4	6.1	153751	2	AC116551	AC116551 Dictyoste
42	72.2	6.1	146098	8	AC105459	AC105459 Homo sapi
c 43	72.2	6.1	250743	2	AE014836	AE014836 Plasmodiu
44	71.8	6.1	110000	14	PFMAL7P1_12	Continuation (13 o
c 45	71.8	6.1	257757	2	AE014837	AE014837 Plasmodiu

ALIGNMENTS

RESULT 1
AF061251
LOCUS AF061251 14002 bp DNA linear BCT 20-AUG-1998
DEFINITION Escherichia coli serotype O157:H7 O antigen gene cluster.
ACCESSION AF061251
VERSION AF061251.1 GI:3435170
KEYWORDS .
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 14002)
AUTHORS Wang,L. and Reeves,P.R.
TITLE Organization of Escherichia coli O157 O antigen gene cluster and identification of its specific genes
JOURNAL Infect. Immun. 66 (8), 3545-3551 (1998)
PUBMED 9673232
REFERENCE 2 (bases 1 to 14002)
AUTHORS Wang,L.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1998) Microbiology, The University of Sydney, NSW 2006, Australia
FEATURES Location/Qualifiers
source 1. .14002
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="C664-1992"
/serotype="O157:H7"
/db_xref="taxon:562"
gene 60. .842
/gene="wbdN"
CDS 60. .842
/gene="wbdN"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase WbdN"
/protein_id="AAC32339.1"
/db_xref="GI:3435171"
/translation="MNKETVSIIMPVYNGAKTIISSVESIIHQSYQDFVLYIIDDCST
DDTFSLINSRYKNNQKIRILRNKTNLGVAESRNYGIEMATGKYISFCDADDLWHEKKL
ERQIEVLNNECVDVVCSNYYVIDNNRNIVGEVNAPHVINYRKMLMKNYIGNLTGIYNA

```
                NKLGKFYQKKIGHEDYLMWLEIINKTNGAICIQDNLAYYMRSNNSLSGNKIKAAKWTW
                SIYREHLHLSFPKTLYYFLLYASNGVMKKITHSLLRRKETKK"
gene            839. .2023
                /gene="wzy"
CDS             839. .2023
                /gene="wzy"
                /codon_start=1
                /transl_table=11
                /product="O antigen polymerase Wzy"
                /protein_id="AAC32340.1"
                /db_xref="GI:3435172"
                /translation="MKSAAKLIFLFLFTLYSLQLYGVIIDDRITNFDTKVLTSIIIIF
                QIFFVLLFYLTIINERKQQKKFIVNWELKLILVFLFVTIEIAAVVLFLKEGIPIFDDD
                PGGAKLRIAEGNGLYIRYIKYFGNIVVFALIILYDEHKFKQRTIIFVYFTTIALFGYR
                SELVLLILQYILITNILSKDNRNPKIKRIIGYFLLVGVVCSLFYLSLGQDGEQNDSYN
                NMLRIINRLTIEQVEGVPYVVSESIKNDFFPTPELEKELKAIINRIQGIKHQDLFYGE
                RLHKQVFGDMGANFLSVTTYGAELLVFFGFLCVFIIPLGIYIPFYLLKRMKKTHSSIN
                CAFYSYIIMILLQYLVAGNASAFFFGPFLSVLIMCTPLILLHDTLKRLSRNENISYNC
                DL"
gene            1992. .2738
                /gene="wbdO"
CDS             1992. .2738
                /gene="wbdO"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase WbdO"
                /protein_id="AAC32341.1"
                /db_xref="GI:3435173"
                /translation="MKISVITVTYNNAEGLEKTLSSLSILKIKPFEIIIVDGGSTDGT
                NRVISRFTSMNITHVYEKDEGIYDAMNKGRMLAKGDLIHYLNAGDSVIGDIYKNIKEP
                CLIKVGLFENDKLLGFSSITHSNTGYCHQGVIFPKNHSEYDLRYKICADYKLIQEVFP
                EGLRSLSLITSGYVKYDMGGVSSKKRILRDKELAKIMFEKNKKNLIKFIPISIIKILF
                PERLRRVLRKMQYICLTLFFMKNSSPYDNE"
gene            2725. .4116
                /gene="wzx"
CDS             2725. .4116
                /gene="wzx"
                /codon_start=1
                /transl_table=11
                /product="O antigen flippase Wzx"
                /protein_id="AAC32342.1"
                /db_xref="GI:3435174"
                /translation="MIMNKIKKILKFCTLKKYDTSSALGREQERYRIISLSVISSLIS
                KILSLLSLILTVSLTLPYLGQERFGVWMTITSLGAALTFLDLGIGNALTNRIAHSFAC
                GKNLKMSRQISGGLTLLAGLSFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKV
                FVIIFGLGIYSNGVQKVYMGIQKAYISNIVNAIFILLSIITLVISSKLHAGLPVLIVS
                TLGIQYISGIYLTINLIIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSG
                DNFIISITLGVTYVAVFSITQRLFQISTVPLTIYNIPLWAAYADAHARNDTQFIKKTL
                RTSLKIVGISSFLLAFILVVFGSEVVNIWTEGKIQVPRTFIIAYALWSVIDAFSNTFA
                SFLNGLNIVKQQMLAVVTLILIAIPAKYIIVSHFGLTVMLYCFIFIYIVNYFIWYKCS
                FKKHIDRQLNIRG"
gene            4113. .5213
                /gene="per"
CDS             4113. .5213
                /gene="per"
                /codon_start=1
                /transl_table=11
                /product="perosamine synthetase Per"
                /protein_id="AAC32343.1"
                /db_xref="GI:3435175"
                /translation="MKMKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKF
                AEQNHVQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFV
                DSDNETWQMSVSDIEQKITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEA
                FGSKYKGKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHR
                QYWHDVIGYNYRMTNICAAIGLAQLEQADDFISRKREIADIYKKNINSLVQVHKESKD
                VFHTYWMVSILTRTAEEREELRNHLADKLIETRPVFYPVHTMPMYSEKYQKHPIAEDL
                GWRGINLPSFPSLSNEQVIYICESINEFYSDK"
gene            5238. .6452
                /gene="wbdP"
CDS             5238. .6452
                /gene="wbdP"
                /codon_start=1
                /transl_table=11
                /product="putative glycosyl transferase WbdP"
                /protein_id="AAC32344.1"
                /db_xref="GI:3435176"
                /translation="MKIALNSDGFYEWGGGIDFIKYILSILETKPEICIDILLPRNDI
                HSLIREKAFPFKSILKAILKRERPRWISLNRFNEQYYRDAFTQNNIETNLTFIKSKSS
                AFYSYFDSSDCDVILPCMRVPSGNLNKKAWIGYIYDFQHCYYPSFFSKREIDQRNVFF
                KLMLNCANNIIVNAHSVITDANKYVGNYSAKLHSLPFSPCPQLKWFADYSGNIAKYNI
                DKDYFIICNQFWKHKDHATAFRAFKIYTEYNPDVYLVCTGATQDYRFPGYFNELMVLA
                KKLGIESKIKILGHIPKLEQIELIKNCIAVIQPTLFEGGPGGGVTFDAIALGKKVILS
                DIDVNKEVNCGDVYFFQAKNHYSLNDAMVKADESKIFYEPTTLIELGLKRRNACADFL
                LDVVKQEIESRS"
gene            6472. .7590
                /gene="gmd"
CDS             6472. .7590
                /gene="gmd"
                /codon_start=1
                /transl_table=11
                /product="GDP-D-mannose dehydratase Gmd"
                /protein_id="AAC32345.1"
                /db_xref="GI:3435177"
                /translation="MTKVALITGVTGQDGSYLAEFLLDKGYEVHGIKRRASSFNTERI
                DHIYQDPHGSNPNFHLHYGDLTDSSNLTRILKEVQPDEVYNLAAMSHVAVSFESPEYT
                ADVDAIGTLRLLEAIRFLGLENKTRFYQASTSELYGLVQEIPQKESTPFYPRSPYAVA
                KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRGLANIAQGLESCLYL
                GNMDSLRDWGHAKDYVRMQWLMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLGIKMSF
                VGKGIEEKGIVDSVEGQDAPGVKPGDVIVAVDPRYFRPAEVDTLLGDPSKANLKLGWR
                PEITLAEMISEMVAKDLEAAKKHSLLKSHGFSVSLALE"
gene            7587. .8558
                /gene="fcl"
CDS             7587. .8558
                /gene="fcl"
                /codon_start=1
                /transl_table=11
                /product="fucose synthetase Fcl"
                /protein_id="AAC32346.1"
                /db_xref="GI:3435178"
                /translation="MMMNKQRIFIAGHQGMVGSAITRRLKQRDDVELVLRTRDELNLL
                DSSAVLDFFSSQKIDQVYLAAAKVGGILANSSYPADFIYENIMIEANVIHAAHKNNVN
                KLLFLGSSCIYPKLAHQPIMEDELLQGKLEPTNEPYAIAKIAGIKLCESYNRQFGRDY
                RSVMPTNLYGPNDNFHPSNSHVIPALLRRFHDAVENNSPNVVVWGSGTPKREFLHVDD
                MASASIYVMEMPYDIWQKNTKVMLSHINIGTGIDCTICELAETIAKVVGYKGHITFDT
                TKPDGAPRKLLDVTLLHQLGWNHKITLHKGLENTYNWFLENQLQYRG"
gene            8561. .9070
                /gene="wbdQ"
CDS             8561. .9070
                /gene="wbdQ"
                /codon_start=1
                /transl_table=11
                /product="WbdQ"
                /protein_id="AAC32347.1"
                /db_xref="GI:3435179"
                /translation="MFLHSQDFATIVRSTPLISIDLIVENEFGEILLGKRINRPAQGY
                WFVPGGRVLKDEKLQTAFERLTEIELGIRLPLSVGKFYGIWQHFYEDNSMGGDFSTHY
                IVIAFLLKLQPNILKLPKSQHNAYCWLSRAKLINDDDVHYNCRAYFNNKTNDAIGLDN
                KDIICLMRQ"
gene            9052. .10500
                /gene="manC"
CDS             9052. .10500
                /gene="manC"
                /codon_start=1
                /transl_table=11
                /product="mannose-1-P guanosyltransferase ManC"
                /protein_id="AAC32348.1"
                /db_xref="GI:3435180"
                /translation="MSDAPIIAVVMAGGTGSRLWPLSRELYPKQFLQLSGDNTLLQTT
                LLRLSGLSCQKPLVITNEQHRFVVAEQLREINKLNGNIILEPCGRNTAPAIAISAFHA
                LKRNPQEDPLLLVLAADHVIAKESVFCDAIKNATPIANQGKIVTFGIIPEYAETGYGY
                IERGELSVPLQGHENTGFYYVNKFVEKPNRETAELYMTSGNHYWNSGIFMFKASVYLE
                ELRKFRPDIYNVCEQVASSSYIDLDFIRLSKEQFQDCPAESIDFAVMEKTEKCVVCPV
                DIGWSDVGSWQSLWDISLKSKTGDVCKGDILTYDTKNNYIYSESALVAAIGIEDMVIV
                QTKDAVLVSKKSDVQHVKKIVEMLKLQQRTEYISHREVFRPWGKFDSIDQGERYKVKK
                IIVKPGEGLSLRMHHHRSEHWIVLSGTAKVTLGDKTKLVTANESIYIPLGAAYSLENP
                GIIPLNLIEVSSGDYLGEDDIIRQKERYKHED"
```

```
     gene            10504. .11874
                     /gene="manB"
     CDS             10504. .11874
                     /gene="manB"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphomannomutase ManB"
                     /protein_id="AAC32349.1"
                     /db_xref="GI:3435181"
                     /translation="MKSLTCFKAYDIRGKLGEELNEDIAWRIGRAYGEFLKPKTIVLG
                     GDVRLTSEALKLALAKGLQDAGVDVLDIGMSGTEEIYFATFHLGVDGGIEVTASHNPM
                     DYNGMKLVREGARPISGDTGLRDVQRLAEANDFPPVDETKRGRYQQINLRDAYVDHLF
                     GYINVKNLTPLKLVINSGNGAAGPVVDAIEARFKALGAPVELIKVHNTPDGNFPNGIP
                     NPLLPECRDDTRNAVIKHGADMGIAFDGDFDRCFLFDEKGQFIEGYYIVGLLAEAFLE
                     KNPGAKIIHDPRLSWNTVDVVTAAGGTPVMSKTGHAFIKERMRKEDAIYGGEMSAHHY

  Query Match            100.0%;  Score 1185;  DB 1;  Length 14002;
  Best Local Similarity  100.0%;  Pred. No. 3.4e-160;
  Matches 1185;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 898

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 958

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 1018

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 1078

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 1138

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 1198

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 1258

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 1318

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 1378

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 1438

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 1498

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 1558

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 1618

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1619 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 1678

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 1738

Qy        901 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 1798

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1858

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1918

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1919 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1978

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1979 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 2023


RESULT 2
BD085555
LOCUS       BD085555               14024 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085555
VERSION     BD085555.1  GI:22631165
KEYWORDS    JP 2001522248-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 14024)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 2 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Escherichia coli
            PN   JP 2001522248-A/2
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .14024
            FT                   /organism='Escherichia coli'.
FEATURES             Location/Qualifiers
     source          1. .14024
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match            100.0%;  Score 1185;  DB 6;  Length 14024;
  Best Local Similarity  100.0%;  Pred. No. 3.4e-160;
  Matches 1185;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 917

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        918 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 977

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 1037

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 1097

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 1157

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 1217

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 1277

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 1337

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 1397

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 1457

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 1517

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 1577

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1578 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 1637

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1638 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 1697

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1698 CAAGTATTTGGAGACATGGGAGCAAATTTTTTTATCAGTTACTACGTATGGAGCAGAACTG 1757

Qy        901 TTAGTTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1758 TTAGTTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 1817

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1818 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1877

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1878 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1937

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1938 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1997

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||||||||||||||||||||||||||||||||||||||
Db       1998 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 2042
```

```
RESULT 3
AX029262
LOCUS       AX029262               14029 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 2 from Patent WO9850531.
ACCESSION   AX029262
VERSION     AX029262.1  GI:10190113
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 2 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .14029
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
ORIGIN

  Query Match           100.0%;  Score 1185;  DB 6;  Length 14029;
  Best Local Similarity  100.0%;  Pred. No. 3.4e-160;
  Matches 1185;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 917

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 977

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 1037

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 1097

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 1157

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 1217

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 1277

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 1337

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 1397

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 1457

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 1517
```

```
Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 1577

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1578 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 1637

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1638 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 1697

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1698 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 1757

Qy        901 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1758 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 1817

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1818 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1877

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1878 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1937

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1938 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1997

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||||||||||||||||||||||||||||||||||||||
Db       1998 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 2042
```

RESULT 4
BD184789/c
LOCUS BD184789 14181 bp DNA linear PAT 17-JUN-2003
DEFINITION Nucleic acid molecule and polypeptide specific to intestinal hemorrhagic pathogenic Escherichia coli O157:H7, and method of use thereof.
ACCESSION BD184789
VERSION BD184789.1 GI:31876989
KEYWORDS JP 2002355074-A/25.
SOURCE Escherichia coli O157:H7
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 14181)
AUTHORS Hayashi,H., Shinagawa,H., Makino,K., Hayashi,T., Onishi,S., Hattori,M. and Kurokawa,K.
TITLE Nucleic acid molecule and polypeptide specific to intestinal hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
JOURNAL Patent: JP 2002355074-A 25 10-DEC-2002;
PRESIDENT OF UNIVERSITY OF TSUKUBA
COMMENT OS Escherichia coli O157:H7
PN JP 2002355074-A/25
PD 10-DEC-2002
PF 24-JAN-2002 JP 2002015959
PI HIDEO HAYASHI,HIDEO SHINAGAWA,KOZO MAKINO,TETSUYA HAYASHI,SHIN
PI ONISHI,
PI MASAHIRA HATTORI,KEN KUROKAWA
PC C12N15/09,C12N15/09,A61K31/7088,A61K39/00,A61K48/00,A61P31/04,
PC C07K14/245,
PC C07K16/12,C12M1/00,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/
PC 02,C12Q1/68,
PC G01N33/15,G01N33/50,G01N33/53,G01N33/53,G01N33/566,G01N37/00,
PC C12N15/00,
PC C12N15/00,C12N5/00
CC Nucleic acid molecule and polypeptide specific to intestinal
CC hemorrhagic
CC pathogenic Escherichia coli O157:H7, and method of use thereof
FH Key Location/Qualifiers
FT source 1. .14181
FT /organism='Escherichia coli O157:H7'.
FEATURES Location/Qualifiers
source 1. .14181
/organism="Escherichia coli O157:H7"
/mol_type="genomic DNA"
/db_xref="taxon:83334"
ORIGIN

Query Match 99.9%; Score 1183.4; DB 6; Length 14181;
Best Local Similarity 99.9%; Pred. No. 5.7e-160;
Matches 1184; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13138 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 13079

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13078 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 13019

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13018 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 12959

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12958 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 12899

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12898 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 12839

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12838 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 12779

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12778 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 12719

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12718 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 12659

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12658 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 12599

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12598 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 12539

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12538 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 12479

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      12478 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAAGTGTTCCATATGTTGTTTCT 12419

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12418 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 12359

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12358 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 12299
```

```
Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12298 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 12239

Qy        901 TTAGTTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12238 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 12179

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12178 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 12119

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12118 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 12059

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12058 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 11999

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||||||||||||||||||||||||||||||||||||||
Db      11998 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 11954


RESULT 5
AR204225/c
LOCUS       AR204225               14187 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 121 from patent US 6365723.
ACCESSION   AR204225
VERSION     AR204225.1  GI:21500820
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14187)
  AUTHORS   Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
  TITLE     Sequences of E. coli O157
  JOURNAL   Patent: US 6365723-A 121 02-APR-2002;
FEATURES             Location/Qualifiers
     source          1. .14187
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           99.9%;  Score 1183.4;  DB 6;  Length 14187;
  Best Local Similarity  99.9%;  Pred. No. 5.7e-160;
  Matches 1184;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13143 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 13084

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13083 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 13024

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13023 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 12964

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12963 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 12904

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12903 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 12844

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12843 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 12784

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12783 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 12724

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12723 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 12664

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12663 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 12604

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12603 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 12544

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12543 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 12484

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db      12483 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAAGTGTTCCATATGTTGTTTCT 12424

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12423 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 12364

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12363 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 12304

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12303 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 12244

Qy        901 TTAGTTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12243 TTAGTTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 12184

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12183 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 12124

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12123 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 12064

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12063 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 12004

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||||||||||||||||||||||||||||||||||||||
Db      12003 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 11959


RESULT 6
AR637574/c
LOCUS       AR637574               14187 bp    DNA     linear   PAT 20-APR-2005
DEFINITION  Sequence 121 from patent US 6855814.
ACCESSION   AR637574
VERSION     AR637574.1  GI:62771316
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 14187)
  AUTHORS   Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
  TITLE     Sequences of E. coli O157
  JOURNAL   Patent: US 6855814-A 121 15-FEB-2005;
            Wisconsin Alumni Research Foundation; Madison, WI
FEATURES             Location/Qualifiers
```

```
     source          1. .14187
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             99.9%;  Score 1183.4;  DB 6;  Length 14187;
  Best Local Similarity   99.9%;  Pred. No. 5.7e-160;
  Matches 1184;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13143 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 13084

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13083 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 13024

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13023 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 12964

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12963 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 12904

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12903 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 12844

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12843 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 12784

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12783 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 12724

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12723 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 12664

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12663 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 12604

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12603 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 12544

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12543 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 12484

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      12483 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAAGTGTTCCATATGTTGTTTCT 12424

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12423 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 12364

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12363 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 12304

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12303 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 12244

Qy        901 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12243 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 12184

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12183 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 12124

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12123 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 12064

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12063 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 12004

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              |||||||||||||||||||||||||||||||||||||||||||||
Db      12003 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 11959


RESULT 7
AB008676
LOCUS       AB008676               31482 bp    DNA     linear   BCT 19-MAY-1999
DEFINITION  Escherichia coli 0157 DNA, map position at 46 min., complete cds.
ACCESSION   AB008676
VERSION     AB008676.1  GI:4867913
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Shimizu,T., Yamasaki,S., Tsukamoto,T. and Takeda,Y.
  TITLE     Analysis of the genes responsible for the O-antigen synthesis in
            enterohaemorrhagic Escherichia coli O157
  JOURNAL   Microb. Pathog. 26 (5), 235-247 (1999)
   PUBMED   10222209
REFERENCE   2  (bases 1 to 31482)
  AUTHORS   Shimizu,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1997) Takeshi Shimizu, International Medical
            Center of Japan, Department of Appropriate Technology Development,
            Research Institute; 1-21-1 Toyama, Shinjuku-ku, Tokyo 162, Japan
            (E-mail:kshimizu@imcj.go.jp, Tel:+81-3-3202-7181,
            Fax:+81-3-3202-7364)
FEATURES             Location/Qualifiers
     source          1. .31482
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="184"
                     /serotype="O157"
                     /db_xref="taxon:562"
                     /map="46 min."
                     /clone="1-4-1"
     gene            505. .1398
                     /gene="galF"
     CDS             505. .1398
                     /gene="galF"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="UTP-glucose-1-phosphate uridylyltransferase"
                     /protein_id="BAA77723.1"
                     /db_xref="GI:4867914"
                     /translation="MTNLKAVIPVAGLGMHMLPATKAIPKEMLPIVDKPMIQYIVDEI
                     VAAGIKEILLVTHASKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNV
                     RQGEPLGLGHSILCARPAIGDNPFVVVLPDVVIDDASADPLRYNLAAMIARFNETGRS
                     QVLAKRMPGDLSEYSVIQTKEPLDREGKVSRIVEFIEKPDQPQTLDSDIMAVGRYVLS
                     ADIWPELERTQPGAWGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKY
                     GLRNLKEGAKFRKGIEKLLSE"
     gene            1717. .2499
                     /gene="wbhA"
     CDS             1717. .2499
                     /gene="wbhA"
                     /note="putative"
                     /codon_start=1
```

/transl_table=11
/product="glycosyl transferase"
/protein_id="BAA77724.1"
/db_xref="GI:4867915"
/translation="MNKETVSIIMPVYNGAKTIISSVESIIHQSYQDFVLYIIDDCST
DDTFSLINSRYKNNQKIRILRNKTNLGVAESRNYGIEMATGKYISFCDADDLWHEKKL
ERQIEVLNNECVDVVCSNYYVIDNKRNIVGEVNAPHVINYRKMLMKNYIGNLTGIYNA
NKLGKFYQKKIGHEDYLMWLEIINKTNGAICIQDNLAYYMRSNNSLSGNKIKAAKWTW
SIYREHLHLSFPKTLYYFLLYASNGVMKKITHSLLRRKETKK"
gene 3153. .3680
/gene="wbhB"
CDS 3153. .3680
/gene="wbhB"
/note="putative"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase"
/protein_id="BAA77725.1"
/db_xref="GI:4867916"
/translation="MLRIINRLTIEQVESVPYVVSESIKNDFFPTPELEKELKAIINR
IQGIKHQDLFYGERLHKQVFGDMGANFLSVTTYGAELLVFFGFLCVFIIPLGIYIPFY
LLKRMKKTHSSINCAFYSYIIMILLQYLVAGNASAFFFGPFLSVLIMCTPLILLHDTL
KRLSRNENISYNCDL"
gene 3649. .4395
/gene="wbhC"
CDS 3649. .4395
/gene="wbhC"
/note="putative"
/codon_start=1
/transl_table=11
/product="glycosyl transferase"
/protein_id="BAA77726.1"
/db_xref="GI:4867917"
/translation="MKISVITVTYNNAEGLEKTLSSLSILKIKPFEIIIVDGGSTDGT
NRVISRFTSMNITHVYEKDEGIYDAMNKGRMLAKGDLIHYLNAGDSVIGDIYKNIKEP
CLIKVGLFENDKLLGFSSITHSNTGYCHQGVIFPKNHSEYDLRYKICADYKLIQEVFP
EGLRSLSLITSGYVKYDMGGVSSKKRILRDKELAKIMFEKNKKNLIKFIPISIIKILF
PERLRRVLRKMQYICLTLFFMKNSSPYDNE"
gene 4388. .5773
/gene="wzx"
CDS 4388. .5773
/gene="wzx"
/note="putative"
/codon_start=1
/transl_table=11
/product="flippase"
/protein_id="BAA77727.1"
/db_xref="GI:4867918"
/translation="MNKIKKILKFCTLKKYDTSSALGREQERYRIISLSVISSLISKI
LSLLSLILTVSLTLPYLGQERFGVWMTITSLGAALTFLDLGIGNALTNRIAHSFACGK
NLKMSRQISGGLTLLAGLSFVITAICYITSGMIDWQLVIKGINENVYAELQHSIKVFV
IIFGLGIYSNGVQKVYMGIQKAYISNIVNAIFILLSIITLVISSKLHAGLPVLIVSTL
GIQYISGIYLTINLIIKRLIKFTKVNIHAKREAPYLILNGFFFFILQLGTLATWSGDN
FIISITLGVTYVAVFSITQRLFQISTVPLTIYNIPLWAAYADAHARNDTQFIKKTLRT
SLKIVGISSFLLAFILVVFGSEVVNIWTEGKIQVPRTFIIAYALWSVIDAFSNTFASF
LNGLNIVKQQMLAVVTLILIAIPAKYIIVSHFGLTVMLYCFIFIYIVNYFIWYKCSFK
KHIDRQLNIRG"
gene 5776. .6870
/gene="wbhD"
CDS 5776. .6870
/gene="wbhD"
/note="putative"
/codon_start=1
/transl_table=11
/product="perosamine synthetase"
/protein_id="BAA77728.1"
/db_xref="GI:4867919"
/translation="MKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKFAE
QNHVQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFVDS
DNETWQMSVSDIEQKITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEAFG
SKYKGKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHRQY
WHDVIGYNYRMTNICAAIGLAQLEQADDFISRKREIADIYKKNINSLVQVHKESKDVF
HTYWMVSILTRTAEEREELRNHLADKLIETRPVFYPVHTMPMYSEKYQKHPIAEDLGW
RGINLPSFPSLSNEQVIYICESINEFYSDK"
gene 6895. .8109
/gene="wbhE"
CDS 6895. .8109
/gene="wbhE"
/note="putative"
/codon_start=1
/transl_table=11
/product="glycosyl transferase"
/protein_id="BAA77729.1"
/db_xref="GI:4867920"
/translation="MKIALNSDGFYEWGGGIDFIKYILSILETKPEICIDILLPRNDI
HSLIREKAFPFKSILKAILKRERPRWISLNRFNEQYYRDAFTQNNIETNLTFIKSKSS
AFYSYFDSSDCDVILPCMRVPSGNLNKKAWIGYIYDFQHCYYPSFFSKREIDQRNVFF
KLMLNCANNIIVNAHSVITDANKYVGNYSAKLHSLPFSPCPQLKWFADYSGNIAKYNI
DKDYFIICNQFWKHKDHATAFRAFKIYTEYNPDVYLVCTGATQDYRFPGYFNELMVLA
KKLGIESKIKILGHIPKLEQIELIKNCIAVIQPTLFEGGPGGGVTFDAIALGKKVILS
DIDVNKEVNCGDVYFFQAKNHYSLNDAMVKADESKIFYEPTTLIELGLKRRNACADFL
LDVVKQEIESRS"
gene 8129. .9247
/gene="gmd"
CDS 8129. .9247
/gene="gmd"
/note="putative"
/codon_start=1
/transl_table=11
/product="GDP-mannose 4,6-dehydratase"
/protein_id="BAA77730.1"
/db_xref="GI:4867921"
/translation="MTKVALITGVTGQDGSYLAEFLLDKGYEVHGIKRRASSFNTERI
DHIYQDPHGSNPNFHLHYGDLTDSSNLTRILKEVQPDEVYNLAAMSHVAVSFESPEYT
ADVDAIGTLRLLEAIRFLGLENKTRFYQASTSELYGLVQEIPQKESTPFYPRSPYAVA
KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRGLANIAQGLESCLYL
GNMDSLRDWGHAKDYVRMQWLMLQQEQPEDFVIATGVQYSVRQFVEMAAAQLGIKMSF
VGKGIEEKGIVDSVEGQDAPGVKPGDVIVAVDPRYFRPAEVDTLLGDPSKANLKLGWR
PEITLAEMISEMVAKDLEAAKKHSLLKSHGFSVSLALE"
gene 9250. .10215
/gene="wbhF"
CDS 9250. .10215
/gene="wbhF"
/note="putative"
/codon_start=1
/transl_table=11
/product="GDP-L-fucose pathway enzyme"
/protein_id="BAA77731.1"
/db_xref="GI:4867922"
/translation="MNKQRIFIAGHQGMVGSAITRRLKQRDDVELVLRTRDELNLLDS
SAVLDFFSSQKIDQVYLAAAKVGGILANSSYPADFIYENIMIEANVIHAAHKNNVNKL
LFLGSSCIYPKLAHQPIMEDELLQGKLEPTNEPYAIAKIAGIKLCESYNRQFGRDYRS
VMPTNLYGPNDNFHPSNSHVIPALLRRFHDAVENNSPNVVVWGSGTPKREFLHVDDMA
SASIYVMEMPYDIWQKNTKVMLSHINIGTGIDCTICELAETIAKVVGYKGHITFDTTK
PDGAPRKLLDVTLLHQLGWNHKITLHKGLENTYNWFLENQLQYRG"
gene 10218. .10727
/gene="wbhG"
CDS 10218. .10727
/gene="wbhG"
/note="putative"
/codon_start=1
/transl_table=11
/product="GDP-L-fucose pathway enzyme"
/protein_id="BAA77732.1"
/db_xref="GI:4867923"
/translation="MFLHSQDFATIVRSTPLISIDLIVENEFGEILLGKRINRPAQGY
WFVPGGRVLKDEKLQTAFERLTEIELGIRLPLSVGKFYGIWQHFYEDNSMGGDFSTHY
IVIAFLLKLQPNILKLPKSQHNAYCWLSRAKLINDDDVHYNCRAYFNNKTNDAIGLDN
KDIICLMRQ"
gene 10739. .12157
/gene="manC"
CDS 10739. .12157
/gene="manC"
/note="putative"
/codon_start=1
/transl_table=11

```
  Query Match             99.9%;  Score 1183.4;  DB 1;  Length 31482;
  Best Local Similarity   99.9%;  Pred. No. 4.2e-160;
  Matches 1184;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2496 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 2555

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2556 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 2615

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2616 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 2675

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2676 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 2735

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2736 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 2795

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2796 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 2855

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2856 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 2915

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2916 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 2975

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2976 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 3035

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3036 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 3095

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3096 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 3155

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       3156 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAAGTGTTCCATATGTTGTTTCT 3215

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3216 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 3275

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3276 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 3335

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3336 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 3395

Qy        901 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3396 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 3455

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3456 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 3515

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3516 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 3575

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3576 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 3635

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||||||||||||||||||||||||||||||||||||||
Db       3636 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 3680
```

```
RESULT 8
AE005174_28/c
WPCOMMENT
  Sequence split into 56 fragments  LOCUS AE005174 Accession AE005174
     Fragment Name      Begin     End
     AE005174_00            1    110000
     AE005174_01       100001    210000
     AE005174_02       200001    310000
     AE005174_03       300001    410000
     AE005174_04       400001    510000
     AE005174_05       500001    610000
     AE005174_06       600001    710000
     AE005174_07       700001    810000
     AE005174_08       800001    910000
     AE005174_09       900001   1010000
     AE005174_10      1000001   1110000
     AE005174_11      1100001   1210000
     AE005174_12      1200001   1310000
     AE005174_13      1300001   1410000
     AE005174_14      1400001   1510000
     AE005174_15      1500001   1610000
     AE005174_16      1600001   1710000
     AE005174_17      1700001   1810000
     AE005174_18      1800001   1910000
     AE005174_19      1900001   2010000
     AE005174_20      2000001   2110000
     AE005174_21      2100001   2210000
     AE005174_22      2200001   2310000
     AE005174_23      2300001   2410000
     AE005174_24      2400001   2510000
     AE005174_25      2500001   2610000
     AE005174_26      2600001   2710000
     AE005174_27      2700001   2810000
     AE005174_28      2800001   2910000
     AE005174_29      2900001   3010000
     AE005174_30      3000001   3110000
     AE005174_31      3100001   3210000
     AE005174_32      3200001   3310000
     AE005174_33      3300001   3410000
     AE005174_34      3400001   3510000
     AE005174_35      3500001   3610000
     AE005174_36      3600001   3710000
     AE005174_37      3700001   3810000
     AE005174_38      3800001   3910000
     AE005174_39      3900001   4010000
     AE005174_40      4000001   4110000
     AE005174_41      4100001   4210000
     AE005174_42      4200001   4310000
     AE005174_43      4300001   4410000
     AE005174_44      4400001   4510000
     AE005174_45      4500001   4610000
     AE005174_46      4600001   4710000
     AE005174_47      4700001   4810000
     AE005174_48      4800001   4910000
     AE005174_49      4900001   5010000
     AE005174_50      5000001   5110000
     AE005174_51      5100001   5210000
     AE005174_52      5200001   5310000
     AE005174_53      5300001   5410000
     AE005174_54      5400001   5510000
```

```
    AE005174_55       5500001   5528445
Continuation (29 of 56) of AE005174 from base 2800001 (AE005174 Escherichia coli O157:H7

  Query Match             99.9%;  Score 1183.4;  DB 1;  Length 110000;
  Best Local Similarity   99.9%;  Pred. No. 2.6e-160;
  Matches 1184;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56709 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 56650

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56649 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 56590

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56589 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 56530

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56529 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 56470

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56469 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 56410

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56409 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 56350

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56349 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 56290

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56289 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 56230

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56229 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 56170

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56169 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 56110

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56109 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 56050

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      56049 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAAGTGTTCCATATGTTGTTTCT 55990

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55989 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 55930

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55929 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 55870

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55869 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 55810

Qy        901 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55809 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 55750

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      55749 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 55690

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55689 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 55630

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55629 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 55570

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||||||||||||||||||||||||||||||||||||||
Db      55569 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 55525


RESULT 9
BA000007_27/c
WPCOMMENT
  Sequence split into 55 fragments  LOCUS BA000007 Accession BA000007
     Fragment Name      Begin      End
     BA000007_00            1   110000
     BA000007_01       100001   210000
     BA000007_02       200001   310000
     BA000007_03       300001   410000
     BA000007_04       400001   510000
     BA000007_05       500001   610000
     BA000007_06       600001   710000
     BA000007_07       700001   810000
     BA000007_08       800001   910000
     BA000007_09       900001  1010000
     BA000007_10      1000001  1110000
     BA000007_11      1100001  1210000
     BA000007_12      1200001  1310000
     BA000007_13      1300001  1410000
     BA000007_14      1400001  1510000
     BA000007_15      1500001  1610000
     BA000007_16      1600001  1710000
     BA000007_17      1700001  1810000
     BA000007_18      1800001  1910000
     BA000007_19      1900001  2010000
     BA000007_20      2000001  2110000
     BA000007_21      2100001  2210000
     BA000007_22      2200001  2310000
     BA000007_23      2300001  2410000
     BA000007_24      2400001  2510000
     BA000007_25      2500001  2610000
     BA000007_26      2600001  2710000
     BA000007_27      2700001  2810000
     BA000007_28      2800001  2910000
     BA000007_29      2900001  3010000
     BA000007_30      3000001  3110000
     BA000007_31      3100001  3210000
     BA000007_32      3200001  3310000
     BA000007_33      3300001  3410000
     BA000007_34      3400001  3510000
     BA000007_35      3500001  3610000
     BA000007_36      3600001  3710000
     BA000007_37      3700001  3810000
     BA000007_38      3800001  3910000
     BA000007_39      3900001  4010000
     BA000007_40      4000001  4110000
     BA000007_41      4100001  4210000
     BA000007_42      4200001  4310000
     BA000007_43      4300001  4410000
     BA000007_44      4400001  4510000
     BA000007_45      4500001  4610000
     BA000007_46      4600001  4710000
     BA000007_47      4700001  4810000
     BA000007_48      4800001  4910000
     BA000007_49      4900001  5010000
     BA000007_50      5000001  5110000
     BA000007_51      5100001  5210000
     BA000007_52      5200001  5310000
```

```
     BA000007_53       5300001   5410000
     BA000007_54       5400001   5498450
Continuation (28 of 55) of BA000007 from base 2700001 (BA000007 Escherichia coli O157:H7

  Query Match             99.9%;  Score 1183.4;  DB 1;  Length 110000;
  Best Local Similarity   99.9%;  Pred. No. 2.6e-160;
  Matches 1184;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86495 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 86436

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86435 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 86376

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86375 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 86316

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86315 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 86256

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86255 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 86196

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86195 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 86136

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86135 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 86076

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86075 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 86016

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86015 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 85956

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85955 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 85896

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85895 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 85836

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      85835 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAAGTGTTCCATATGTTGTTTCT 85776

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85775 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 85716

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85715 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 85656

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85655 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 85596

Qy        901 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85595 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 85536

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85535 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 85476

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85475 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 85416

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85415 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 85356

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||||||||||||||||||||||||||||||||||||||
Db      85355 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 85311


RESULT 10
AY730593
LOCUS       AY730593               14323 bp    DNA     linear   BCT 25-OCT-2004
DEFINITION  Salmonella enterica subsp. enterica serovar Urbana UDP-glucose
            pyrophosphorylase (galF) gene, partial cds; O-antigen gene cluster,
            complete sequence; and 6-phosphogluconate dehydrogenase (gnd) gene,
            partial cds.
ACCESSION   AY730593 AY012194
VERSION     AY730593.1  GI:54401420
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Urbana
  ORGANISM  Salmonella enterica subsp. enterica serovar Urbana
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 14323)
  AUTHORS   Jensen,S.O. and Reeves,P.R.
  TITLE     Molecular evolution of the GDP-mannose pathway genes (manB and
            manC) in Salmonella enterica
  JOURNAL   Microbiology (Reading, Engl.) 147 (Pt 3), 599-610 (2001)
   PUBMED   11238967
REFERENCE   2  (bases 1 to 14323)
  AUTHORS   Samuel,G., Hogbin,J.P., Wang,L. and Reeves,P.R.
  TITLE     Relationships of the Escherichia coli O157, O111, and O55 O-Antigen
            Gene Clusters with Those of Salmonella enterica and Citrobacter
            freundii, Which Express Identical O Antigens
  JOURNAL   J. Bacteriol. 186 (19), 6536-6543 (2004)
   PUBMED   15375135
REFERENCE   3  (bases 1 to 14323)
  AUTHORS   Jenkins,S.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-2000) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
REFERENCE   4  (bases 1 to 14323)
  AUTHORS   Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2004) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
  REMARK    Sequence update by submitter
COMMENT     On Oct 25, 2004 this sequence version replaced gi:11992707.
FEATURES             Location/Qualifiers
     source          1. .14323
                     /organism="Salmonella enterica subsp. enterica serovar
                     Urbana"
                     /mol_type="genomic DNA"
                     /strain="M284"
                     /serotype="30"
                     /serovar="Urbana"
                     /sub_species="enterica"
                     /db_xref="taxon:260677"
     gene            <1..852
                     /gene="galF"
     CDS             <1. .852
                     /gene="galF"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-glucose pyrophosphorylase"
                     /protein_id="AAV34505.1"
```

/db_xref="GI:54401421"
/translation="MHMLPATKAIPKEMLPIVDKPMIQYIVDEIVAAGIKEIVLVTHA
SKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNVRQAQPLGLGHSILC
ARPVVGDNPFIVVLPDIIIDDATADPLRYNLAAMVARFNETGRSQVLAKRMKGDLSEY
SVIQTKEPLDNEGKVSRIVEFIEKPDQPQTLDSDLMAVGRYVLSADIWAELERTEPGA
WGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKYGLRNLKEGAKFRKS
IEQLLHE"

gene 1220. .2002
/gene="wbdN"

CDS 1220. .2002
/gene="wbdN"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAV34506.1"
/db_xref="GI:54401422"
/translation="MNREIVSVIMPVYNGSHTIVDSIESVLNQTYKDIKLYVIDDCST
DDTAKLLTVNYANNKNVKIIISSKNKGVAASRNLGIKESKGKYIAFCDADDLWENNKI
EKQINLLEHNNINVVCSNYYIMDFNKKIIGETHSPVTISYKKMLMKNYVGNLTGVYNA
EKLGKFYQKKVGHEDYLMWLEIIKKSNFAICLQENLAFYRVSNNSLSASKLVAAKWTW
NIYRECLGLSLPKSLYYFSLYAANGVIKKIGLINSDNSKAKE"

gene 2255. .3157
/gene="wzy"

CDS 2255. .3157
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O-antigen polymerase"
/protein_id="AAV34507.1"
/db_xref="GI:54401423"
/translation="MEIMAIILFAKEGIPIFDADPGGAKLKIADGNGIYIRYIKYFGN
IVTFTLLLLLNKQKVKQKIIAIIYFLTIALFGYRSELVLLFLQYYIIMSIVNEGGNKY
KFLNKIFIFIFMGGVIGSLFYFSLGQDGAVNNSTENIDRIINRLTIEQVESVPYIISE
SLEYDLFPTPELGKEINAIFNRILGEKDQNLFYGERLHKKVFGDMGANFLSVTTYGAE
LLAFFGPLCIFLIPIGLCIPFAMLKKISNTHSTVNGAIYSFLIMILLQYLIAGNASAF
FFGPFITAIVMSIPLMILHDILKR"

gene 3163. .3894
/gene="wbdO"

CDS 3163. .3894
/gene="wbdO"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAV34508.1"
/db_xref="GI:54401424"
/translation="MNTSVITVTYNNAEGLNKTLESLCLLKVKPYEIIIIDANSIDNT
KAIVKKYSDMNIIYICEKDNGIYDAMNKGLRLARGKLLHYLNAGDCVSGEIYADLNGP
CLLKVGLFDKDNLIGFSSLTHSNTGYCHQGIIFPSKHPIYDLKYKLCADYKLIQMVFP
MGLKQLPLISNGFIKYDMGGVSANKRFLRDKELLQIMLETNKKTLIKFIPLMVIKMVF
PKKIRRILRRVNYYFSLILTNKKNV"

gene 3896. .5287
/gene="wzx"

CDS 3896. .5287
/gene="wzx"
/codon_start=1
/transl_table=11
/product="O-antigen flippase"
/protein_id="AAV34509.1"
/db_xref="GI:54401425"
/translation="MIKEKIKILFKYCTLRKYNTTTDYGRQQERYRIISLSIISNFIS
KLLALLSLILTVSLTLPYLGQDRFGIWMTIASFSAALTFLDLGIGNALTNRIAHSFAR
GKKIRMIRQISGGLTLLAILAFIIIAIFLTIANFVNWGSLLKGINDNTHDEVQFAIKL
FAIIFGIGIFSNGVQRIYMGMQKAYISYISNAFFILISIISLFISSRLHAGIPMLILS
TIGVQYLSGLSLTILLIYKKLITLNKIRINVKKETPYLISNGFFFFILQMGTLATWSG
DNFIISTTLGVTYVSILSITQRLFQISTVPLTIYNAPLWAAYADAHARHDVDFIRKTL
LGSLKIVGISSTTLALILVIFGQSIIAIWTGEELRVPQSLIVAYAVWSIIDAYSNTLS
SFLNGLNIIKQQVFAVVTLILISIPSKYFIVSHFGLTAMLSCFIILYVTNYLIWYKCS
FKKSIHLQLNIRG"

gene 5290. .6384
/gene="per"

CDS 5290. .6384
/gene="per"
/codon_start=1
/transl_table=11
/product="perosamine synthetase"
/protein_id="AAV34510.1"
/db_xref="GI:54401426"
/translation="MEYIPVYQPALSGKEKEYVNDCIDSTWISSKGSYIQRFEKKFAE
QNNIQYATTVSNGTVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPVFVDS
DLQTWQISVDDIKNKITSRTKAIMCVHLYGHPCDMSPILEIAKVNNLFIIEDCAEAFG
SKYKDKFVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAAHRQY
WHDVIGYNYRMTNICAAIGLAQLEQAQLFIARKREIAEIYKNNLTPLIEVHQEGNDIF
HTYWMVSILARNAEEREVLRSFLSEKLTETRPLFYPVHTMPMYSDKYQKHPVAENLGW
RGINLPSFPTLSNEQVIYICNAINEFYKDK"

gene 6409. .7623
/gene="wbdP"

CDS 6409. .7623
/gene="wbdP"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAV34511.1"
/db_xref="GI:54401427"
/translation="MKITLNTDGFNEWGGGVDFIKYLLRTLEIEQDIFLDVLLPREDI
SSLLKRNIFPVKNMIKAILSGQTPRWKKLEGFNEQYYKHAFAEITEQSKLSFITGKSS
AYFSHYIKSDYDIILPCMRTPAVNMKNRAWLGYIYDFQHCYYPSFFSEREIAKRNLFF
KRMLDNANNVIVNSHAVISDAKKFVGDFSANLHALPFSPCPELNWLNDNSDLIDKYAI
DKDYFIICNQFWKHKDHSTAFKAFKKYIEINPDVYLVCTGATQDYRFPGYFDELMCLL
KKLGISEKVKILGHIPKLEQIELIKNSIAVIQPTLFEGGPGGGITFDAVALGKKIVLS
DIEVNKDVDCGHVYFFHSQNEYSLTEALVKATAVDHYYDPHVLLESGIKRRKKCADFI
LNVISKEIESRI"

gene 7647. .8765
/gene="gmd"

CDS 7647. .8765
/gene="gmd"
/codon_start=1
/transl_table=11
/product="GDP-mannose dehydratase"
/protein_id="AAV34512.1"
/db_xref="GI:54401428"
/translation="MSKVALITGVTGQDGSYLAEFLLEKGYEVHGIKRRASSFNTERV
DHIYQDPHSCNPKFHLHYGDLTDASNLTRILQEVQPDEVYNLGAMSHVAVSFESPEYT
ADVDAMGTLRLLEAIRFLGLEKKTRFYQASTSELYGLVQEIPQKETTPFYPRSPYAVA
KLYAYWITVNYRESYGIYACNGILFNHESPRRGETFVTRKITRAIANIAQGLESCLYL
GNMDSLRDWGHAKDYVRMQWMMLQQEQPEDFVIATGVQYSVRQFVELAAAQLGIKLRF
EGEGINEKGIVVSVTGHDAPGVKPGDVIVAVDPRYFRPAEVETLLGDPSKAHEKLGWK
PEITLSEMVSEMVANDLEAAKKHSLLKSHGFNVNLSLE"

gene 8768. .9733
/gene="fcl"

CDS 8768. .9733
/gene="fcl"
/codon_start=1
/transl_table=11
/product="fucose synthetase"
/protein_id="AAV34513.1"
/db_xref="GI:54401429"
/translation="MNKQRIFVAGHRGMVGSAIVRQLAQRGDVELVLRTRDELDLLDG
RAVQAFFAGTGIDQVYLAAAKVGGIVANNTYPADFIYENMMIESNIIHAAHLHNVNKL
LFLGSSCIYPKLARQPMAESELLQGTLEPTNEPYAIAKIAGIKLCESYNRQYGRDYRS
VMPTNLYGPHDNFHPDNSHVIPALLRRFHEAAQSHAPEVVVWGSGTPMREFLHVDDMA
AASIHVMELSREVWQENTAPMLSHINVGTGVDCTISELAQTIAKVVGYQGRVVFDAAK
PDGTPRKLLDVTRLHQLGWYHEISLEAGLAGTYQWFLENQQRFRG"

gene 9736. .10236
/gene="gmm"

CDS 9736. .10236
/gene="gmm"
/codon_start=1
/transl_table=11
/product="GDP-mannose mannosylhydrolase"
/protein_id="AAV34514.1"
/db_xref="GI:54401430"
/translation="MFLRQEDFAAVVRATPLISLDFIVENGQGEILLGQRLNRPAQGY
WFVPGGRVRKDETLEAAFARLTQAELGVRLPLAAGTFYGVWQHFYDDNFSGEDFSTHY
IVLGFRLRVAESDLLLPDAQHGSYRWLTPEQLLASDNVHDNSRAYFQNAPYSVIGLDK
KDVKYV"

```
  Query Match             44.0%;  Score 521.6;  DB 1;  Length 14323;
  Best Local Similarity   66.6%;  Pred. No. 1.3e-65;
  Matches  793;  Conservative    0;  Mismatches  389;  Indels    8;  Gaps    3;

Qy          2 TGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGT 61
              ||||   | | | |||  || ||| | | || ||| |  || | |||  ||| ||||| |
Db       2007 TGAACAGATCTGTTAAACTGGTTTGTCTCTTTCTACTATCATTCTATGCTCTTCAGTTAT 2066

Qy         62 ATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTA 121
               |||  |   | ||||   | || | | ||   ||| ||   |  | ||||||  |
Db       2067 TTGGAATAGCAACAGATATAAGGATTAAAGATGACAATATAAGAATCATTACTAGTTCTC 2126

Qy        122 TAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAA 181
              | ||  |||  ||  || ||||| | ||||||||| | |||    ||  ||  |||||
Db       2127 TTATCCTATACCAAGTTATTTTTATCTTATTATTTTTTATAATATATAGGAAAAAAAGAC 2186

Qy        182 AACAGCAGA---AAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTT 238
              || | || |   || | || ||  | |||    |  | ||  || || || |||| ||||
Db       2187 AAGAACAAACACAATATTTCATTATAAACACTAAAATTAAAGTAGTATTCTTTTTACTTT 2246

Qy        239 TTGTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTG 298
              || | ||  | ||||||   ||| || || | |||   ||||| || || ||||| ||||
Db       2247 TTATAACCGTGGAAATTATGGCTATAATTCTCTTTGCCAAAGAGGGAATACCTATTTTTG 2306

Qy        299 ATGATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGAT 358
              |||  || || || ||||| || |||| ||| || || || |||||| | || || ||||
Db       2307 ATGCGGACCCTGGAGGGGCAAAGCTTAAAATTGCGGATGGCAATGGAATATATATAAGAT 2366

Qy        359 ATATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATA 418
              |||| || ||||||||||||||||||    ||  | ||| |  | ||      || | || |
Db       2367 ATATAAAATATTTTGGTAATATAGTAACCTTCACTTTATTGTTACTCCTAAATAAACAAA 2426

Qy        419 AATTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATC 478
              || | |||||  |  |  ||   | |  |||||||   ||| |||||| ||||||| ||||
Db       2427 AAGTAAAACAAAAAATAATAGCAATAATATATTTCTTAACAATTGCTCTATTTGGCTATC 2486

Qy        479 GTTCTGAATTGGTGTTGCTCATTCTTCAA---TATATATTGATTACCAATATCCTGTCAA 535
              | || || |||||  || |  | ||||||   ||||||  | ||  | | |  |      |
Db       2487 GATCAGAGTTGGTACTGTTATTCCTTCAATATTATATAATTATGTCAATTGTTAATGAAG 2546

Qy        536 AGGATAACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTG 595
               || |||      ||       |||| |  ||  |    | | ||||  |||  || ||
Db       2547 GGGGTAATAAATATAAATTTTTAAATAAGATTTTCATTTTTATTTTTATGGGGGGAGTCA 2606

Qy        596 TATGCTCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATA 655
              |  |    || |||||| | |||||||||||| || || |    |||| |      | | |
Db       2607 TTGGTAGTTTATTTTATTTTAGTTTAGGACAGGATGGTGCGGTAAATAATAGTACTGAAA 2666

Qy        656 ATATGTTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTG 715
              ||||    || || ||||||||  |||| || ||||||||||||| ||||||||||  ||
Db       2667 ATATTGATAGAATCATTAATAGACTAACTATTGAGCAAGTTGAAAGTGTTCCATATATTA 2726

Qy        716 TTTCTGAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAG 775
              ||||||| ||   |  |||   | || || |||||||||||  |  || ||| |||| |
Db       2727 TTTCTGAGTCCTTAGAATACGACCTATTCCCAACACCAGAGTTGGGTAAAGAAATAAACG 2786

Qy        776 CAATAATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTAC 835
              | ||  | ||||||||    |||   ||  | ||| | |||||||||||  |||||  |||
Db       2787 CTATCTTTAATAGAATTTTAGGAGAGAAAGACCAAAATTTATTTTATGGCGAACGACTAC 2846

Qy        836 ATAAACAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAG 895
              ||||| |||| |||||||| ||||| || |||||||||||||| |||||  ||||||||||
Db       2847 ATAAAAAAGTTTTTGGAGATATGGGGGCTAATTTTTTATCAGTCACTACATATGGAGCAG 2906

Qy        896 AACTGTTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATAC 955
              |||| || | |||||||||    | | ||  ||||  | || ||  ||||  |||  |||||
Db       2907 AACTTTTGGCTTTTTTTGGGCCTTTATGCATATTTTTAATTCCCATAGGATTATGTATAC 2966

Qy        956 CTTTTTATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATT 1015
              |||||    | || || | |||   ||| || |||||| |  | ||| | ||  | ||||
Db       2967 CTTTTGCAATGTTGAAAAAAATATCAAATACACATAGCACTGTTAATGGTGCTATATATT 3026

Qy       1016 CATATATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTT 1075
              ||| | | |||||||||||||| ||||| ||| | ||||| |||||||| || |||||||
Db       3027 CATTTTTAATTATGATTTTATTACAATATTTAATTGCTGGAAATGCATCTGCTTTCTTTT 3086

Qy       1076 TTGGTCCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGT 1135
              ||||||| ||| |  | | | |  ||||| |||  ||  ||||  || | ||||| |  |
Db       3087 TTGGTCCATTTATTACAGCAATTGTAATGAGTATACCATTGATGATACTACATGACATAT 3146

Qy       1136 TAAAGAGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||| ||    |  ||||| | ||| |||||||| ||||| || ||
Db       3147 TAAAGAGATAAT--AGTTATGAATACATCTGTTATAACAGTGACATACAA 3194


RESULT 11
AY730592
LOCUS       AY730592               14224 bp    DNA     linear   BCT 02-NOV-2004
DEFINITION  Citrobacter freundii serotype 90 UDP-glucose pyrophosphorylase
            (galF) gene, partial cds; O-antigen gene cluster, complete
            sequence; and 6-phosphogluconate dehydrogenase (gnd) gene, partial
            cds.
ACCESSION   AY730592
VERSION     AY730592.2  GI:55228555
KEYWORDS    .
SOURCE      Citrobacter freundii
  ORGANISM  Citrobacter freundii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1  (bases 1 to 14224)
  AUTHORS   Samuel,G., Hogbin,J.P., Wang,L. and Reeves,P.R.
  TITLE     Relationships of the Escherichia coli O157, O111, and O55 O-Antigen
            Gene Clusters with Those of Salmonella enterica and Citrobacter
            freundii, Which Express Identical O Antigens
  JOURNAL   J. Bacteriol. 186 (19), 6536-6543 (2004)
   PUBMED   15375135
REFERENCE   2  (bases 1 to 14224)
  AUTHORS   Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2004) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
REFERENCE   3  (bases 1 to 14224)
  AUTHORS   Reeves,P.R., Samuel,G., Hogbin,J.-P. and Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2004) School of Molecular and Microbial
            Biosciences, University of Sydney, Sydney, NSW 2006, Australia
  REMARK    Sequence update by submitter
COMMENT     On Nov 2, 2004 this sequence version replaced gi:54401406.
FEATURES             Location/Qualifiers
     source          1. .14224
                     /organism="Citrobacter freundii"
                     /mol_type="genomic DNA"
                     /strain="M1972"
                     /serotype="90"
                     /db_xref="taxon:546"
     gene            <1. .773
                     /gene="galF"
     CDS             <1. .773
                     /gene="galF"
                     /codon_start=3
                     /transl_table=11
                     /product="UDP-glucose pyrophosphorylase"
                     /protein_id="AAV34492.2"
                     /db_xref="GI:55228556"
                     /translation="FVIVAAGIKEIVLVTHASKNAVENHFDTSYELESLLEQRVKRQL
                     LAEVQSICPPGVTIMNVRQAQLLGLGHSILCARPIIGDNPFVVVLPDVVLDTASADPL
                     RYNLAAMVARFNETGRSQVLAKRMEGDLSEYSVIQTKDPLDNEGKVSRIVEFIEKPDQ
                     PQTLDSDLMAVGRYVLSADIWAELEKTAPGAWGRIQLTDAISEPAKKQSVDAILMTGE
                     SYDCGKKMGYMQAFVKYGLRNLKEGPKFRKSIEKLLAM"
     gene            1104. .1853
                     /gene="wbdN"
     CDS             1104. .1853
                     /gene="wbdN"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
```

/protein_id="AAV34493.1"
/db_xref="GI:54401408"
/translation="METESVSIIMPVYNGASTIVQSIESILNQTYKDIRLYIIDDCST
DETSILLQNLYISNDRVELIRNIKNKGVAESRNIGIEQSRGKYIAFCDADDLWGETKL
EKQLDILESEKVDVVCTNYYLIDLEDNIIGSVCSPGIITYKTMLGKNHIGNLTGIYNV
KKIGKVYQKKIGHEDYLMWLEILRKTEQAICIQENLAKYRVSKKSLSANKLIAAQWTW
NIYRKQLKIPIIPALYFFSQYAVQGLLKKII"

gene 1890. .3047
/gene="wzy"

CDS 1890. .3047
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O-antigen polymerase"
/protein_id="AAV34494.1"
/db_xref="GI:54401409"
/translation="MKKSVGIISFFLFSFYVLQLYGLLIDVRIKDVNIQIIAGGLVLF
QILCVLFYALFIRKNLSYQSGLFFNSKKIIIIFLLTFVTIELAAVILFLKEGIPIFDA
DPGGAKLRIADGNGIYIRYIKYFGNIVCFSLCLLLKDRKILHYSVVIIYFSTISLFGY
RSELIMLMIQYYILMSIINNNKVKHSMLRKVVILFSISCIVIGLFYFSLGQEGQENNY
ASNVNRIFNRLTIEQVESVPYIINESMKYDLFPTSEIGKEINAIANRFQGIKDNSLFY
GERLHKDIFGDMGENFLSVTTFGAELLAFFGPFCFFLIPLGIYIPYFLLIKMSQTKNT
VLCSAYSYSVVIFLQYLIAGNFSAFFFGPILTIVIMCIPLVLLHDVLKRLN"

gene 3047. .3781
/gene="wbdO"

CDS 3047. .3781
/gene="wbdO"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAV34495.1"
/db_xref="GI:54401410"
/translation="MKISVVTVTYNNGNGLRKTLQSLSSIKTKPCEIIIIDAESKDNT
KSVVDEYFSKMNIIYINEKDNGIYDAMNKGHKIANGDYIHYLNAGDCVSGDVYANVTG
TCLIKVGLFNERDSFLGFSNLTHSNTGYCHQGVIFPKEHLDYDLKYKLCADYDLLQKT
FPEGLASLPLITDGYIKYDMSGVSANKTLLRDKELLAIMWKNNKRIIPKFMIYVLFKT
MIPKGCRRFLRKLKYSISVFSLNGNK"

gene 3789. .5180
/gene="wzx"

CDS 3789. .5180
/gene="wzx"
/codon_start=1
/transl_table=11
/product="O-antigen flippase"
/protein_id="AAV34496.1"
/db_xref="GI:54401411"
/translation="MNNLLRKIYSVLKYCTLKKFDTSTDYGRQQERYRIISLSIIANL
ISKFLSLLTLVLTVSLTLPYLGPARFGVWMTIASLSAALTFLDLGIGNALTNRIAHAC
ANGERTEVIKKTSGGIALLLLVSICITCVLYIICSVIDWRIFIKGINDNILGEVQRAI
KLFVVIFGISILSNGIQKIYMGMQKAYITYFFNSLLSLISIFCLIYYSSLHATIPVLI
LISSGIPLFTGMLLMILLIKDGFVSLNEIFHNMRNEAPHLISTGFFFFILQIGTLATW
SGDNFIISSTLGVAFVAILSLTQRLFQISTVPLTIYNTPLWVAYADAHARKDTHFIKK
TLLASFKIVGFSSLIMATFLILFGRDIISVWTGDKIDIPQSFIIAYAIWSIIDAVSNT
LASFLNGLGIVKQQMFAVVSLVIISIPAKYFIVSHFGLTAMLSCFIAIYIVNYIVWYT
ISFKKCIYSQLSI"

gene 5188. .6285
/gene="per"

CDS 5188. .6285
/gene="per"
/codon_start=1
/transl_table=11
/product="perosamine synthetase"
/protein_id="AAV34497.1"
/db_xref="GI:54401412"
/translation="MKYIPVYEPSLTGKEKEYVNDCLDSTWISSKGSYIQRFETEFSK
HISAQYATTVSNGTVAIHLALIALGISEGDEVIVPTLTYIASVNAINYTGAIPVFVDS
DAETWQVSIEDIRNKITNKTKAIMCVHLYGQPCDMDAILEIANEFNLFVIEDCAEAFG
SKYKDKYVGTFGDISTFSFFGNKTITTGEGGMVVTNDKTLYERCVHYKGQGLAAHRQY
WHDVIGYNYRMTNICAAIGLAQLEQAEGFINRKRNIAEIYKKELFNSPVKVHQESDNT
FHTYWMVSILTRTAEEREELRNYLSENFIETRPLFYPVHTMPMYSEKYQKHPVAEDLG
WRGMNLPSYPGMTDEQVIYICNKIKCFYNEK"

gene 6310. .7524
/gene="wbdP"

CDS 6310. .7524
/gene="wbdP"
/codon_start=1
/transl_table=11
/product="glycosyltransferase"
/protein_id="AAV34498.2"
/db_xref="GI:55228557"
/translation="MKITINTDGFNEWGGGVDFIKYLLLILSTKDDVFTDIIVPKDDF
ASLFRRNVFPIKNAIKSLGKGKFPRWVKKDGFDEGYYHNAFSGLVDKSKLTFINSNSN
SYFDYHAKSDNDIIFPCMRVPTNSVCDSPWIGYIYDFQHCYYPSFFSAREIKKRELYF
REMLSSANNVIVNANSVIADAKNFIGDFSADLHALPFSPCPQQIWLDDNSNLVDKYAI
SKDYFIVCNQFWKHKNHVVVFKAFKEYIKVNPEAYLVCTGGIQDYRFPEYFGELKALI
QKLGISSKIKILGHIPKLEQVELIKKSIAVIQPTLFEGGPGGGATYDAVALGKKVILS
DIDVNREVNCGDVIFFNPQDEYSLAKALFKAEEEKSVKENGDLMKLGLKRREKCADFI
MSVIRKEIESRN"

gene 7546. .8664
/gene="gmd"

CDS 7546. .8664
/gene="gmd"
/codon_start=1
/transl_table=11
/product="GDP-mannose dehydratase"
/protein_id="AAV34499.1"
/db_xref="GI:54401414"
/translation="MSKVALITGVTGQDGSYLAEFLLEKGYEVHGIKRRASSFNTGRV
DHIYQDPHLGNPKFNLHYGDLTDTSNLTRILKEVQPDEVYNLGAMSHVAVSFESPEYT
ADVDAMGTLRLLEAIRFLGLEKKTRFYQASTSELYGLVREIPQKETTPFYPRSPYAVA
KLYAYWITVNYRESYGMYACNGILFNHESPRRGETFVTRKITRAMANIAQGLESCLYL
GNMDSLRDWGHAKDYVRMQWMMLQQEQPEDFVIATGVQYSVRDFVEMSAAQLGIKLSF
EGDGVNEKGIVASVTGHDAPGVKVGDVIVEVDPRYFRPAEVETLLGDPSKAHEKLGWK
PEITLQEMVAEMVASDLEAAKKYSLLKESGFDVNLSLE"

gene 8667. .9632
/gene="fcl"

CDS 8667. .9632
/gene="fcl"
/codon_start=1
/transl_table=11
/product="fucose synthetase"
/protein_id="AAV34500.1"
/db_xref="GI:54401415"
/translation="MKKQRIFVAGHRGMVGTAIVRQLSLRDNVELVLRTRDELNLLDS
SAVQAFFATERIDQVYLAAAKVGGIVANNTYPADFIYENMMIESNIIHAAHLHNVNKL
LFLGSSCIYPKQATQPIAESELLQGTLEPTNEPYAIAKIAGIKLCESYNRQYGRDYRS
VMPTNLYGPHDNFHPSNSHVIPALLRRFHEAREQNAPDVVVWGSGTPMREFLHVDDMA
AASIHVMELDSEIWQEYTQPMLSHINVGTGVDCTIRELAQTIAQVVGYKGKVVFDSSK
PDGTPRKLLDVTRLHSLGWRHSVSLEYGLESTYQWFLENQHWCRG"

gene 9635. .10135
/gene="gmm"

CDS 9635. .10135
/gene="gmm"
/codon_start=1
/transl_table=11
/product="GDP-mannose mannosylhydrolase"
/protein_id="AAV34501.1"
/db_xref="GI:54401416"
/translation="MFLHHDEFATVVRSTPLVSIDLIVENADGEFLLGKRLNRPAQGY
WFVPGGRVQKDERIENAFERLTLAELGLRLPLSVAEFYGVWQHFYDDNFSGPDFSTHY
IVLAFHIRGQESCLCLPDEQHKIWRWFKSEELEEEKDVHVNSRAYFHIENSNVIGLDN
KGLNHV"

gene 10128. .11576
/gene="manC"

CDS 10128. .11576
/gene="manC"
/codon_start=1
/transl_table=11
/product="mannose-1-P guanosyltransferase"
/protein_id="AAV34502.1"
/db_xref="GI:54401417"

Query Match 38.1%; Score 451.2; DB 1; Length 14224;
Best Local Similarity 62.9%; Pred. No. 1.5e-55;
Matches 749; Conservative 0; Mismatches 433; Indels 8; Gaps 3;

```
Qy          2 TGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGT 61
              ||||    | | |     | | ||   ||| ||  | ||| |  |||||  |||||||||| |
Db       1891 TGAAAAAATCTGTAGGGATAATAAGTTTTTTTTTGTTTTCTTTTTATGTTCTCCAGTTAT 1950

Qy         62 ATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTA 121
              |||||  |   | ||  |||| ||| || | ||  || ||| | |   |  || |  |
Db       1951 ATGGGCTCTTAATTGATGTTCGAATTAAAGATGTTAATATACAAATTATAGCCGGAGGAT 2010

Qy        122 TAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAA 181
              || || |||| || || || | |||| | || | |  | ||   |||  |||  |    |
Db       2011 TAGTTTTATTCCAAATATTATGTGTTCTCTTCTATGCTTTATTTATTCGAAAAAATTTAT 2070

Qy        182 AACAGCAGAAAAAATTTATCGTGAACT---GGGAGCTAAAGTTAATACTCGTTTTCCTTT 238
              | | ||||     ||  |  | || |      |  |||   | ||| || | ||    |
Db       2071 CATATCAGAGTGGTTTGTTTTTTAATTCAAAAAAAATAATCATTATATTCCTATTGACAT 2130

Qy        239 TTGTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTG 298
              |||| || || ||  | || |||||| ||||||| ||||| ||||||||||||||| ||||
Db       2131 TTGTAACGATTGAGCTCGCAGCTGTAATTTTATTCCTTAAGGAAGGTATTCCTATTTTTG 2190

Qy        299 ATGATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGAT 358
              |||  ||||| || ||||| ||||| || || || || || |||||  | || |||||||
Db       2191 ATGCAGATCCGGGAGGGGCGAAACTGAGGATTGCAGATGGGAATGGGATCTATATTAGAT 2250

Qy        359 ATATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATA 418
              |||| || |||||||| ||||| |||   |||  |||  | |||| |     | || |  |
Db       2251 ATATAAAATATTTTGGAAATATTGTTTGCTTTAGTTTATGTCTTCTCCTAAAAGATCGAA 2310

Qy        419 AATTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATC 478
              || |    ||      |  |  |  | |||||||| |||| ||   ||| || || ||||
Db       2311 AAATACTGCACTATTCTGTTGTTATAATATATTTTTCAACTATAAGTTTGTTCGGCTATC 2370

Qy        479 GTTCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGG 538
              ||||||||||  |  || | ||  |||| |||    |  | |    |||  |    ||
Db       2371 GTTCTGAATTAATCATGTTAATGATTCAGTATTATATTCTAATGTCTATTATTAATAATA 2430

Qy        539 ATAACCGTAATCCTAAAATAAAAAGAATAATAGGG---TATTTTTTATTGGTAGGGGTTG 595
              ||||    || |  |  ||   |||||   | | |   |  ||||   |     |  |||
Db       2431 ATAAGGTCAAGCACAGCATGCTAAGAAAGGTGGTGATCTTATTTTCTATTAGTTGTATTG 2490

Qy        596 TATGCTCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATA 655
              |       || |||||  | |||||||| || || ||  || ||||| | |      |  |
Db       2491 TTATTGGTTTATTTTACTTTAGTTTAGGTCAGGAAGGGCAAGAAAATAATTACGCAAGCA 2550

Qy        656 ATATGTTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTG 715
              || |    |||||  |||||||| | ||||| || || || ||| |||| |||||| ||
Db       2551 ATGTAAATAGGATTTTTAATAGGCTTACAATTGAACAGGTAGAAAGTGTCCCATATATTA 2610

Qy        716 TTTCTGAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAG 775
              ||  ||| || || ||  | ||| | ||||| ||  | ||  ||| ||||||  |||| |
Db       2611 TTAATGAGTCAATGAAATATGATCTATTTCCTACTTCTGAAATAGGAAAGGAGATAAACG 2670

Qy        776 CAATAATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTAC 835
              | ||    |||||| | || ||||| ||  ||  |      | |||||||| || || ||||
Db       2671 CTATTGCTAATAGATTTCAAGGAATTAAAGATAATAGTCTTTTTTATGGCGAGCGTTTAC 2730

Qy        836 ATAAACAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAG 895
              | ||| |  | ||||| |||||||||| ||| || |||||  |||||  || | ||| || |
Db       2731 ACAAAGATATTTTTGGTGACATGGGAGAAAACTTCTTATCTGTTACAACATTTGGGGCTG 2790

Qy        896 AACTGTTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATAC 955
              | || || |  |||||||    | ||||  | ||  | |||||  | || |||||  || |
Db       2791 AGCTTTTGGCCTTTTTTGGCCCTTTCTGCTTTTTTTTGATCCCGCTTGGAATATACATTC 2850

Qy        956 CTTTTTATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATT 1015
              | | || ||| ||||  | ||||  | ||||  | | |   | |    ||  ||   || |
Db       2851 CATATTTTCTCTTAATAAAAATGTCACAAACAAAAAATACAGTTTTATGTTCAGCATACT 2910

Qy       1016 CATATATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTT 1075
              | |||    |  | || || || || ||  || | |||||||||   || || || ||||
Db       2911 CTTATTCAGTGGTTATATTTTTACAGTATCTAATAGCTGGGAATTTTTCTGCATTTTTTT 2970

Qy       1076 TTGGTCCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGT 1135
              |||||||  | ||  |  || | |||||||||| |||| |  | ||| | ||||||   |
Db       2971 TTGGTCCCATCCTAACTATAGTTATAATGTGTATTCCTTTAGTATTACTTCATGATGTCT 3030

Qy       1136 TAAAGAGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              | |||||||||  |   |||||||| |||||  ||||||| ||||||||
Db       3031 TGAAGAGATTA--AATTAATGAAAATTTCAGTAGTAACTGTAACTTATAA 3078
```

RESULT 12
CR847965

```
LOCUS       CR847965              190356 bp    DNA     linear   HTG 08-JUL-2005
DEFINITION  Danio rerio chromosome 24 clone CH211-223F21, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   CR847965
VERSION     CR847965.3  GI:70663878
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 190356)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Jul 8, 2005 this sequence version replaced gi:54034964.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC223F21
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 187632 bases at least Q40
            Consensus quality: 187888 bases at least Q30
            Consensus quality: 188392 bases at least Q20
            Insert size: 189856; sum-of-contigs
            Insert size: 198701; 0.8% error; agarose-fp
            Quality coverage: 8.57x in Q20 bases; sum-of-contigs Quality
            coverage: 8.26x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11266: contig of 11266 bp in length
            *    11267    11366: gap of 100 bp
            *    11367    28114: contig of 16748 bp in length
            *    28115    28214: gap of 100 bp
            *    28215   150363: contig of 122149 bp in length
            *   150364   150463: gap of 100 bp
            *   150464   171430: contig of 20967 bp in length
            *   171431   171530: gap of 100 bp
            *   171531   181870: contig of 10340 bp in length
            *   181871   181970: gap of 100 bp
            *   181971   190356: contig of 8386 bp in length.
FEATURES             Location/Qualifiers
     source          1. .190356
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="24"
                     /clone="CH211-223F21"
```

```
                     /clone_lib="CHORI-211"
     misc_feature    1. .11266
                     /note="assembly_fragment:00213
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    11367. .28114
                     /note="assembly_fragment:00636
                     fragment_chain:1"
     misc_feature    28215. .150363
                     /note="assembly_fragment:00906
                     fragment_chain:1"
     misc_feature    150464. .171430
                     /note="assembly_fragment:00401
                     fragment_chain:1"
     misc_feature    171531. .181870
                     /note="assembly_fragment:00090
                     fragment_chain:1"
     misc_feature    181971. .190356
                     /note="assembly_fragment:00001
                     fragment_chain:1"
ORIGIN
```

Query Match 7.0%; Score 82.8; DB 14; Length 190356;
Best Local Similarity 41.6%; Pred. No. 0.0018;
Matches 430; Conservative 0; Mismatches 592; Indels 12; Gaps 4;

```
Qy        19 TTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATAGAT 78
             ||  ||| |||| |  | ||||  | || |  |||  ||||  |              ||| ||
Db    181644 TTATTTTATTTAATGTTTTTTATGCATTTTGTTCTGTAGTTTAAGCAAAACCCAATAAAT 181703

Qy        79 GATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATT 138
             || || ||| ||| |      ||| ||    |  | ||| || || ||     ||
Db    181704 TTATTTAGAAAAAAATTTTAGCGTTTGGTGTTGTAGACACTCATAGTTTTAGTTATAGTT 181763

Qy       139 TTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTT 198
             | | |  | |||   | ||| | |     |||| ||  |  | | |      | | || |
Db    181764 ATATATTATATATATATATATTATTATATATATATATATATATATATATATATATATAT 181823

Qy       199 ATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCT 258
             ||    |    | |  || |   || |      ||  | |      |
Db    181824 ATATATAGAGTGTATATATATAAAAAAAAAAAATTTTTCTAAAAAAANNNNNNNNNNNNN 181883

Qy       259 GCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCT 318

Db    181884 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 181943

Qy       319 AAACTTAGAATAGCTGAAGGTAATGGACTT---TACATTAGATATATTAAGTATTTTGGT 375
                                           ||   || | |  || ||| || || ||
Db    181944 NNNNNNNNNNNNNNNNNNNNNNNNNNNNATTATATATAATTTATTTATAAAATAATTATTA 182003

Qy       376 AATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACC 435
             |||    || |  |   |||| | |  | | | ||     |||||||    | | |
Db    182004 AATTATTTTTTAATAATATTATATTTATTATTTTATTTTAATAAATTATTATAAATATTT 182063

Qy       436 ATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTG 495
             |   | ||   || | | | |     |   |||       || |  || | |||   | ||
Db    182064 AAATTTTTAAAATTTATAATATAATTAAATTTTAAAAATTTTAATTATAAATATTTATTA 182123

Qy       496 CTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAA 555
             | ||   | |  ||||| | || |  | |||      ||   ||||     |  || |
Db    182124 TTAATATATAATAATATAATAATAAATATTAT------AATTAATAATTTATTTTATATA 182177

Qy       556 ATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTA 615
             | ||    ||||      ||| ||||   ||     |  ||    |  || || ||| |
Db    182178 AATAATTTTTTAATTATAATTTTATTATAATTATTTTTAATAAAATATTTATTATATATT 182237

Qy       616 AGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAATTAAT 675
             | | ||  | |      ||  |      | ||  | ||  || |||    || ||||||
Db    182238 ATTATATTAATATTTTATAAATTTATAATTTTATTAATTATAATATTATATATTATTAAT 182297

Qy       676 AGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCTATTAAGAAC 735
             |    || | || |   | ||  |  | ||  | ||    ||| |  || | |||| ||
```

```
Db    182298 AATAAAATATTAAATATAATTTTAATTTTTTTAAATAAAATTTTTATATATTTTAATAAT 182357

Qy       736 GATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATAATAAATAGAATACAG 795
             ||||  |||      | |  || | |  | ||  ||   ||  | | |     |
Db    182358 TATTTTTTTTTATTTATATAAATATATATATATTTTAATTTATTTTATTTTTTTTTTA 182417

Qy       796 G--GAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAACAAGTATTTGGAG 853
               | || |  |     | | |  ||||     |    |  || |   |   |
Db    182418 TTTAATTATAAAAATTTAATTAAAATTATATTTTATTTATTTATTATTTAAATTAATATT 182477

Qy       854 ACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTGTTAGTTTTTTTTG 913
             | |    |   |  ||||| |   ||| ||   ||  |         | | |  | ||| |
Db    182478 ATAATTTATATATATTTTTTTTTATTATTATTAATTTATTTTTTTTATAATATATTTATA 182537

Qy       914 GTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTTTATCTTTTAAAGA 973
             |||  | | | |||  | ||    ||||    ||||||||  | | ||| | | |||
Db    182538 TATTTATTTTTATATATAATA-ATATTTATTTTTATATATATATATATATTTATAAAAAT 182596

Qy       974 GAATGAAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATATATCATTATGATTT 1033
             ||  | ||| |  |||     || |||      | |||  ||||||   |||   |
Db    182597 TAAATATAAATATTTATATAATATTATATTTATATATATATATATATATATATATATA 182656

Qy      1034 TATTGCAATACTTA 1047
             |||   |||  ||
Db    182657 TATATATATATATA 182670
```

```
RESULT 13
CR628390
LOCUS       CR628390              165165 bp    DNA     linear   HTG 16-JUL-2005
DEFINITION  Danio rerio chromosome 21 clone DKEYP-18E10, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   CR628390
VERSION     CR628390.4  GI:70910566
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 165165)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Jul 16, 2005 this sequence version replaced gi:51174174.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp18E10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 162177 bases at least Q40
            Consensus quality: 162871 bases at least Q30
            Consensus quality: 163348 bases at least Q20
            Insert size: 164565; sum-of-contigs
            Insert size: 179126; 3.2% error; agarose-fp
            Quality coverage: 9.07x in Q20 bases; sum-of-contigs Quality
            coverage: 8.70x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
            * be preserved.
            *        1     5837: contig of 5837 bp in length
            *     5838     5937: gap of 100 bp
            *     5938    39552: contig of 33615 bp in length
            *    39553    39652: gap of 100 bp
            *    39653    42452: contig of 2800 bp in length
            *    42453    42552: gap of 100 bp
            *    42553    64586: contig of 22034 bp in length
            *    64587    64686: gap of 100 bp
            *    64687    92339: contig of 27653 bp in length
            *    92340    92439: gap of 100 bp
            *    92440   129431: contig of 36992 bp in length
            *   129432   129531: gap of 100 bp
            *   129532   165165: contig of 35634 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165165
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="21"
                     /clone="DKEYP-18E10"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .5837
                     /note="assembly_fragment:00036
                     fragment_chain:1"
     misc_feature    5938. .39552
                     /note="assembly_fragment:00870
                     fragment_chain:1"
     misc_feature    39653. .42452
                     /note="assembly_fragment:00004"
     misc_feature    42553. .64586
                     /note="assembly_fragment:00090.0"
     misc_feature    64687. .92339
                     /note="assembly_fragment:00450"
     misc_feature    92440. .129431
                     /note="assembly_fragment:01294"
     misc_feature    129532. .165165
                     /note="assembly_fragment:01774"
ORIGIN

  Query Match             6.8%;  Score 80.6;  DB 14;  Length 165165;
  Best Local Similarity   42.3%;  Pred. No. 0.004;
  Matches  340;  Conservative    0;  Mismatches  457;  Indels    6;  Gaps    2;

Qy         19 TTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATAGAT 78
              | | | || |||||  ||||   | | |    | |  | | |||     |||
Db       5023 TAGTTATTATTATTATTATTATTATTATTATATATATATATATATATATATATATATA 5082

Qy         79 GATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATT 138
               ||      | | || |  ||| | |  |||  |    |||    ||  || | |  ||
Db       5083 TATATATATATATATATATATATATATATATATATATATATATATATTAATATATTTT 5142

Qy        139 TTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTT 198
              |||||| |||| | |||||  | |     ||||   |   | ||   |  | ||||   |
Db       5143 TTTTTTTTTTTNTAATTTTTTTTTTTTTTTATTTTTTTTATAATTAAAAAAAAAANTATA 5202

Qy        199 ATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCT 258
              ||    |      |  |||  |  |     |     | ||| |   | || | ||   |
Db       5203 ATATATATAAAAAAAAAAAATATTTTTTATATAAAAATATTTTTTTTATTTATATATAAAT 5262

Qy        259 GCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCT 318
                 |   | |  | || |   |  | ||  ||||| |    || ||  |
Db       5263 AAATATATAAAAATAATAAATAATTATNTTTATATATATATATATAATATAAAAAAAAAA 5322

Qy        319 AAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAAT 378
              |||   | || |    ||  || |    |||| ||     | |||||   | ||   ||
Db       5323 AAAAAAAAAANAAAAAAANNTATTTANNTTTATATATNNAAAATTAAANNTATTTTAAAA 5382

Qy        379 ATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATC 438
               || |  |  ||   ||  || || ||  |  |    | |||   ||| | |  |
Db       5383 TTATTNTTTNTTTTTTTTNTTTTTTTTATTNTTTNNNAAAAAAAAAAAAAAAAAAA---- 5438

Qy        439 ATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTC 498
                 |||     | | |  | ||        | ||    |     |  | || |    ||  |
Db       5439 ATAAANAAAAAATAAAAAATTNNNNTNTAATANNTTATNNNTATTAAAATTNNTTTTNTT 5498

Qy        499 ATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATA 558
              | | |   ||| |   |  ||           |||  | ||    ||   ||  |||
Db       5499 AATATNNTATAAANTATATTTNNTNNNNNTAAANAAAAAAAAAAAAAAAAAAAATANTATA 5558

Qy        559 AAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGT 618
              |||| | |  |       ||||   |        || |||     || || ||| ||| |
Db       5559 AAAATATTTTTTTTAATNTTTNNNTTTANNTNTTTTTATATNTTTTTTTTATATTTAATT 5618

Qy        619 TTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAATTAATAGG 678
                | | ||         | | ||  | |   ||||        |||    ||  ||
Db       5619 AAATAATAATTATATTTAAATAAAAAATATAATAAANNNTATATAANNNNAAAAAAATTA 5678

Qy        679 TTAACAATAGAGCAAGTTGAAGGTGTT--CCATATGTTGTTTCTGAATCTATTAAGAACG 736
              ||  |  |  || | |  | ||   |||  |  ||    |  ||| || ||
Db       5679 AAAANTAATTATTAAATATTATATATTAATTATAAAATAATTTATTAATTATAAATAAAT 5738

Qy        737 ATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATAATAAATAGAATACAGG 796
              || | ||      |    | |  | ||  |  || ||    |   || ||  ||| |
Db       5739 ATATATTANNTTAAAATTAATATNATAATTANATATAAAANTTTTAAANTAANATATATA 5798

Qy        797 GAATAAAGCATCAAGACTTATTT 819
               | ||||  || |    | | ||
Db       5799 AATTAAATTATTATATATAAATT 5821


RESULT 14
AC105933/c
LOCUS       AC105933              175370 bp    DNA     linear   PRI 03-SEP-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-454I5, complete sequence.
ACCESSION   AC105933 AC108735
VERSION     AC105933.2  GI:22657522
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 175370)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175370)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 175370)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On or before Sep 3, 2002 this sequence version replaced
            gi:19172752, gi:18129410.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: BCM
            -------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-454I5 (bc0430)
            ------------- Summary Statistics
                Sequencing vector: plasmid; 90% of reads
                Sequencing vector: plasmid; L08752; 10% of reads
                Chemistry: Dye-terminator ET; 40% of reads
```

Chemistry: Dye-terminator Big Dye; 60% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 175145 bases at least Q40
Consensus quality: 175352 bases at least Q30
Consensus quality: 175369 bases at least Q20
Insert size: 175370; sum-of-contigs
Quality coverage: 8.3x in Q20 bases; sum-of-contigs

Overlapping Sequences:
5': RP11-776K4 (UWGC:bc0559) AC105750, 50334-bp overlap
3': RP11-464H2 (UWGC:bc0433) AC099053

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

EcoRI		NsiI		BglII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
8696	8947	9493	9693	4142	4059
6	<800	774	<800	2067	2052
5550	5431	1832	1822	7992	8163
1570	1526	5765	5635	4134	4059
247	<800	85	<800	5474	5433
490	<800	15076	14784	831	866
699	<800	6264	6149	2887	2878
4483	4439	1302	1322	884	866
4507	4439	1928	1822	1187	1145
1356	1298	3879	3983	4348	4302
2485	2545	4001	3983	2354	2502
6141	6060	3568	3585	7060	6888
3775	3666	1124	1118	2503	2658
709	<800	1717	1662	9139	8994
1313	1298	315	<800	388	<800
7305	7188	2515	2562	2275	2342
4945	4840	58	<800	992	985
3101	3025	459	<800	617	<800
3619	3666	4509	4381	93	<800
679	<800	2071	2129	221	<800
2153	2177	2786	2864	2593	2658
3483	3666	7223	7609	2809	2781
2139	2177	69	<800	1692	1649
6661	6590	6254	6149	1554	1518
474	<800	120	<800	202	<800
10268	10040	391	<800	432	<800
11153	10780	579	<800	316	<800
3689	3666	1808	1822	1724	1723
8344	8947	14480	14784	3102	3116
1988	1974	7	<800	2186	2211
8027	8222	1516	1511	11787	11021
3730	3666	978	991	11304	11021
3619	3666	2765	2864	1135	1145
5935	6060	796	<800	184	<800
399	<800	1510	1511	2249	2211
2573	2545	653	<800	53	<800
685	<800	4001	3983	1788	1723
3260	3275	3010	3064	282	<800
5813	5760	3205	3228	157	<800
2587	2545	3209	3228	8376	8163
1134	1136	1020	1040	3451	3487
405	<800	1355	1322	606	<800
756	<800	6265	6149	10374	11021
3277	3275	298	<800	4329	4302
301	<800	9959	9693	4088	4059
1794	1770	1347	1322	4594	4517
2535	2545	8260	8242	2675	2658
3293	3275	1517	1511	6798	6888
2413	2545	1104	1118	11189	11021
4035	3920	6057	6149	983	985

4834	4840	1636	1597	3877	3815
941	945	2556	2562	3745	3633
2227	2371	1205	1214	5224	5168
392	<800	6312	6149	4769	4718
5815	5760	2202	2241	21	<800

Query Match 6.8%; Score 80; DB 8; Length 175370;
Best Local Similarity 46.7%; Pred. No. 0.0048;
Matches 423; Conservative 0; Mismatches 475; Indels 8; Gaps 5;

```
Qy        69 TATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTAT 128
             ||| ||| || ||  ||||| | ||||| | |   ||  |||||  || ||| |||
Db     67424 TATTATATATTATAATATAATATATTTTTACA-TAATATATTATATATTATAATATAATA 67366

Qy       129 ATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCA 188
               ||   || |  | | || |||  | | || |||    | |  ||  || | | |
Db     67365 TATTTTTATATAATATTATATATTATAATATAATATATTTTTATATAATATATTATAT 67306

Qy       189 GAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTAT 248
               | | ||   |    |   |   | |||  |    || |  ||||    | |||
Db     67305 ATTATATTATAATATATTTTTATATAATATATTATAATATATTATATTTTTATATAATAT 67246

Qy       249 AGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCC 308
             |  |  |  |  | || ||||||  |  || |  |  |  |  |||||||    || ||
Db     67245 ATTATATTATAATATATTTTTATATAATATATTATATTATAATATATTTTTATATAATAT 67186

Qy       309 AGGGGGGGCTAAACT-TAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGT 367
             |        || | | ||  |||  |  |  ||||  | | || |||| | |||| |  |
Db     67185 ATTATATATTATAATATAATATATTTTATATAATATATTATATATTATA-ATATAATAT 67127

Qy       368 ATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAAC 427
             |||||    | |  |||    ||  ||   ||||     ||| | |   || ||
Db     67126 ATTTTATATACAATGTTTATGTTATATATTTTATATACAATGTTTATGTTATATATTTTA 67067

Qy       428 AGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAAT 487
                 |    |     || |||||||||  |    | || ||   ||   |||   | | |||
Db     67066 TATACAATGTTTATGTTATATATTTTATATACAATGTTTATATTATATATAAATATAAAT 67007

Qy       488 TGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTA 547
               | |   |  | |  |||| ||    ||   |    | |  ||    |
Db     67006 ATATATAAATATATATAATATATAAATATTATATATAATATTTATATAATATATATAAAT 66947

Qy       548 ATCCTAAAATAAAAAGAATA----ATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCG 603
             |||   |||||   | ||||    |||   |||  | |||   ||     |  |||  |
Db     66946 ATCTATAAATATTTATAATATAATATAAAATATAATATATATTTATATATAATATAATAT 66887

Qy       604 TTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTA 663
               |    |    | | | ||  |  | |   | || |      |   |||| | || | |
Db     66886 ATAAATATATTTAATATATAAATATATTTATAATATGTAAATAAATATATTTATTTATAG 66827

Qy       664 AGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAA 723
             |  ||| ||||||  |    |||| |  |  |   |    |      || |  | |   |
Db     66826 AATATACTTAATATATATTAAATATATAATATATAATATAAATATATAATATATTATAAATA 66767

Qy       724 TCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATAATA 783
             | |||||    |  ||  |  | || | | |   || |   | || |      ||||||
Db     66766 TATATTATATATAATACATATTATATACTATATTATCAATATATAATATATTATATAATA 66707

Qy       784 AATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAACAA 843
               |||  ||| |    || |   || |    | ||  |||     |     || | |   |
Db     66706 CATATTATATAATATATTATATATAATATATAATAATATAATTATTATATATAATATATA 66647

Qy       844 GTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTGTTA 903
               ||  |  |   ||  |   ||| | | || |  || |   |||  |  | |  |  ||
Db     66646 ATA-ATATAATTATTATATATAATATATAATAATATAATTATTATATATAATACATAATA 66588

Qy       904 GTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTTTAT 963
```

```
                 | ||||   | || | | |    |   ||     | |   | |||||||   |  |||
Db     66587 TATATTTTATATATTATATATAATATATATAATATATAAAATACATATATAAGATAATAT 66528

Qy         964 CTTTTA 969
                 || ||
Db     66527 ATTATA 66522
```

RESULT 15
AR579680
LOCUS AR579680 1141 bp DNA linear PAT 14-DEC-2004
DEFINITION Sequence 22 from patent US 6784342.
ACCESSION AR579680
VERSION AR579680.1 GI:56583130
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1141)
AUTHORS Kunst,L. and Clemens,S.
TITLE Regulation of embryonic transcription in plants
JOURNAL Patent: US 6784342-A 22 31-AUG-2004;
The University of British Columbia; Vancouver;
CAX;
FEATURES Location/Qualifiers
source 1. .1141
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Query Match 6.5%; Score 77.4; DB 6; Length 1141;
Best Local Similarity 10.6%; Pred. No. 0.079;
Matches 110; Conservative 392; Mismatches 536; Indels 2; Gaps 1;

```
Qy          7 TCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGG 66
              ::::    : : ::|:::::::         ::| : ::  || ::|:  :: :| :: :
Db         67 KYRRWYNNKSRWWKGWYKKKWYBCANNTSBRYHARRWKDMKTAYBMTMTNKWGKTGWRHR 126

Qy         67 GTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATT 126
              : :  ::: :: :::  :  |   :   ::::   ::  :    | ::::|: :::
Db        127 YWRWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNATGWDDDTKYHMWNN 186

Qy        127 ATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAG 186
              :|  : : ::|: :::::   : ::::: ::::   ::::::: :: :::::: :|
Db        187 NGCBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRKVRRWVRTRG 246

Qy        187 CAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACT 246
              ::  ::|::| :  ::    |:  |   : :       | : :  ::  :   :
Db        247 RMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRAKYWGWNRABVNSTC 306

Qy        247 ATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGAT 306
              |::   |:: |:|:       :| :::::: ::|| |  :       :|::  ::: ::
Db        307 TTWKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNWTYKKARHBARW 366

Qy        307 CCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAG 366
              :  :  :: :| | :: :  : ::::    :     |  ::: : |::::::::::
Db        367 DWVWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYN 426

Qy        367 TATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAA 426
                  : :  : : ::  | ::::     :  ::: | |:    | ::: | :   ::
Db        427 NNNNGGRTYYGWTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKTYWNNCYWKS 486

Qy        427 CAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAA 486
              :   ::: :|  :: ::  :: :: :       : :::   |: :::: :::    :  :
Db        487 MTNGKSHRBAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBVCSKWW--NNYAAWY 544

Qy        487 TTGGTGTTGCTCATTCTTCAATATATTGATTACCAATATCCTGTCAAAGGATAACCGT 546
              |:::: :|: :::::   ::::: :|:  |:   : :|::: ::   ::::  : ::::
Db        545 TKSSWNYTSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWSHTWBHBR 604

Qy        547 AATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTG 606
              |   :::: :   |    ::|::  :::|                          ||    :::
Db        605 AGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNNNNNNATCARDDYYA 664
```

```
Qy        607 TTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGG 666
Db        665 ASRWYAMANAKWYYYKBAANNAYYTHANNWWGCWNNATDTRRTMWKNNNNNNAGTWKNNN 724

Qy        667 ATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCT 726
Db        725 NNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADAAABTTDKRNNGAYTKYTTTN 784

Qy        727 ATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATAATAAAT 786
Db        785 NNNTYRGVVTNTAARDGWANNNNNNNNNNNNNNGWSDMWVTWWAYANYGTNNNNNNNNN 844

Qy        787 AGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAACAAGTA 846
Db        845 AYAWWTNKWYYTTDDRWRBAYTNNNNNNRMAYYGAYADDYAYYMSDTCDAWMKWDATKMN 904

Qy        847 TTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTGTTAGTT 906
Db        905 NATTYNRGTAWRTNNNNNNMTMKTKYYBHAAWNNNNNNGKMCTAHTWWVCKATKTTKGCW 964

Qy        907 TTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTTTATCTT 966
Db        965 MNCTTTCRKYKNNCTWYTWMTTTRTTWYAATRWKTNNATGSMTRCNATGWKNNNYWTGWK 1024

Qy        967 TTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATATATCATT 1026
Db       1025 TRWTAYRMATRWMKAWWKVMATGSWNTNSYARWAYKTRAYKGWYYNACAWRWRWGKATCY 1084

Qy       1027 ATGATTTTATTGCAATACTT 1046
Db       1085 MTDNAWWTACATSWMATHKY 1104
```

Search completed: May 10, 2006, 20:34:03
Job time : 4306.31 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 09:42:47 ; Search time 553.175 Seconds
(without alignments)
14276.972 Million cell updates/sec

Title: US-09-423-093B-2_COPY_858_2042
Perfect score: 1185
Sequence: 1 gtgaagtcagcggctaagtt.....gttataactgtgacttataa 1185

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1185	100.0	14024	2	AAX06749	Aax06749 E. coli O
	2	1185	100.0	14024	3	AAZ56386	Escherich
c	3	1183.4	99.9	14181	10	ADC01214	Adc01214 Enterohae
c	4	1183.4	99.9	14187	9	ACD19108	Acd19108 E. coli 0
	5	75.2	6.3	115218	8	ACA64845	Aca64845 Human HNR
c	6	74.8	6.3	115218	8	ACA64845	Aca64845 Human HNR
	7	71	6.0	8056	8	ABZ10100	Abz10100 Haematopo
	8	70	5.9	8056	8	ABZ10246	Abz10246 Haematopo
c	9	69.6	5.9	1554	14	ADZ70999	Adz70999 Human chr
c	10	68.6	5.8	6816	12	ADQ24856	Adq24856 Human sof
	11	68.4	5.8	5728	6	ABL32101	Abl32101 Human imm
c	12	68	5.7	38678	14	AEB32373	Aeb32373 Human gen
c	13	68	5.7	38684	14	AEB32391	Aeb32391 Human gen
	14	67.8	5.7	14551	6	ABL34585	Abl34585 Human met
	15	67.8	5.7	14551	7	ADS99846	Ads99846 Complemen
	16	67.6	5.7	8222	8	ACF62816	Acf62816 Colon can
	17	67.6	5.7	11222	10	ADB54318	Adb54318 Pretreate
	18	67	5.7	15500	4	AAS46508	Aas46508 Tumour su
	19	66.8	5.6	2131	14	ADZ71009	Adz71009 Human chr
	20	66.4	5.6	6255	6	ABL32960	Abl32960 Human imm
	21	66.4	5.6	15732	4	AAS45388	Aas45388 Chemicall
	22	66.4	5.6	15732	6	ABK28233	Abk28233 DNA trans
	23	66.2	5.6	6954	6	ABL33390	Abl33390 Human imm
	24	66	5.6	8222	8	ACF62794	Acf62794 Colon can
	25	66	5.6	11222	10	ADB54190	Adb54190 Pretreate
c	26	65.2	5.5	2131	14	ADZ71009	Adz71009 Human chr
	27	65	5.5	15548	6	ABL34155	Abl34155 Human imm
	28	63.8	5.4	13131	6	ABL92249	Abl92249 Chemicall
	29	63.4	5.4	18683	6	ABL32313	Abl32313 Human imm
	30	63.4	5.4	18683	6	ABL54334	Abl54334 Chemicall
	31	63.4	5.4	46951	10	ADE13891	Ade13891 Human opt
	32	62.8	5.3	745	14	ADZ70917	Adz70917 Human chr
	33	62.4	5.3	9510	4	AAS46437	Aas46437 Tumour su
	34	62.4	5.3	9510	6	ABL34562	Abl34562 Human met
	35	62.4	5.3	9510	7	ADS99823	Ads99823 Bisulphit
	36	62.4	5.3	9859	13	ADS89714	Ads89714 Oligonucl
	37	62.2	5.2	14924	6	ABL32224	Abl32224 Human imm
	38	62.2	5.2	14924	6	ABL54321	Abl54321 Chemicall
c	39	62	5.2	99764	13	ABD32992	Abd32992 Human can
	40	61.8	5.2	5930	6	ABL32517	Abl32517 Human imm
	41	61.6	5.2	6534	4	AAS45404	Aas45404 Chemicall
	42	61.6	5.2	6534	6	ABL33426	Abl33426 Human imm
	43	61.6	5.2	6534	6	ABK28251	Abk28251 DNA trans
	44	61.6	5.2	6534	6	ABN80184	Abn80184 Human che
c	45	61.4	5.2	1864	1	AAN71405	Aan71405 Sequence

ALIGNMENTS

RESULT 1
AAX06749
ID AAX06749 standard; DNA; 14024 BP.
XX
AC AAX06749;
XX
DT 26-APR-1999 (first entry)
XX
DE E. coli O157 antigen gene cluster.
XX
KW O antigen; O157 antigen; diagnosis; wbdN gene; wzx gene; wzy gene;
KW wbdP gene; wbdO gene; wbdR gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 79. .861
FT /*tag= a
FT /product= "sugar transferase"
FT /note= "wbdN gene (ORF1), this region is specifically
FT claimed in Claim 6; encodes AAW88312"
FT CDS 858. .2042
FT /*tag= b
FT /product= "O_antigen-polymerase"
FT /note= "wzy gene (ORF2), this region is specifically
FT claimed in Claim 6; encodes AAW88313"
FT CDS 2011. .2757
FT /*tag= c
FT /product= "sugar_transferase"
FT /note= "wbdO gene (ORF3), this region is specifically
FT claimed in Claim 6; encodes AAW88314"
FT CDS 2744. .4135
FT /*tag= d
FT /product= "O_antigen_flippase"
FT /note= "wzx gene (ORF4), this region is specifically
FT claimed in Claim 6; encodes AAW88315"
FT CDS 4132. .5232
FT /*tag= e
FT /product= "perosamine_synthetase"
FT /note= "per gene (ORF5); encodes AAW88316"
FT CDS 5257. .6471

```
FT                   /*tag=  f
FT                   /product= "sugar_transferase"
FT                   /note= "wbdP gene (ORF6), this region is specifically
FT                   claimed in Claim 6; encodes AAW88317"
FT   CDS             6491. .7609
FT                   /*tag=  g
FT                   /note= "ORF7; encodes AAW88318"
FT   CDS             7606. .8578
FT                   /*tag=  h
FT                   /note= "ORF8; encodes AAW88319"
FT   CDS             8580. .9089
FT                   /*tag=  i
FT                   /note= "ORF9; encodes AAW88320"
FT   CDS             9071. .10519
FT                   /*tag=  j
FT                   /note= "manC gene (ORF10); encodes AAW88321"
FT   CDS             10523. .11893
FT                   /*tag=  k
FT                   /note= "manD gene (ORF11); encodes AAW88322"
FT   repeat_region   12007. .13123
FT                   /tag= l
FT                   /note= "H-repeat, may play a role in gene cluster
FT                   assembly"
FT   CDS             13156. .13821
FT                   /*tag=  m
FT                   /product= "N-acetyl_transferase"
FT                   /note= "wbdR gene (ORF12), this region is specifically
FT                   claimed in Claim 6; encodes AAW88323"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88312, AAW88313, AAW88314, AAW88315, AAW88316, AAW88317,
DR   AAW88318, AAW88319, AAW88320, AAW88321, AAW88322, AAW88323.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 6; Fig 8; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Escherichia coli O157 O antigen. The gene cluster was
CC   obtained by PCR amplification of E. coli O157:H7 strain (C664-1992) DNA
CC   using primers (see AAX06752-63) based on O antigen gene sequences. 12
CC   Open reading frames within the gene cluster were identified on the basis
CC   of homology to known sequences. The genes encode proteins (see AAW88312-
CC   23) involved the synthesis of sugars present in the polysaccharide
CC   antigen, and in the transport or processing of polysaccharide or
CC   oligosaccharide units. The use of nucleic acid molecules derived from
CC   particular assembly and transport genes, particularly wbd (transferase),
CC   wzx (flippase) and wzy (polymerase) genes, within O antigen gene clusters
CC   improves the specificity of methods for the detection and identification
CC   of O antigens, e.g. in tests of food- or faecal- derived samples, or
CC   samples from patients. The O antigen is a major virulence factor of
CC   enteropathogenic E. coli strains that cause diarrhoea and haemorrhagic
CC   colitis
XX
SQ   Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 1185;  DB 2;  Length 14024;
  Best Local Similarity 100.0%;  Pred. No. 4.6e-205;

  Matches 1185;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 917

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 977

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 1037

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 1097

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 1157

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 1217

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 1277

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 1337

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 1397

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 1457

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 1517

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 1577

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1578 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 1637

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1638 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 1697

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1698 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 1757

Qy        901 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1758 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 1817

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1818 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1877

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1878 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1937
```

```
Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1938 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1997

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
             |||||||||||||||||||||||||||||||||||||||||||||
Db       1998 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 2042


RESULT 2
AAZ56386
ID   AAZ56386 standard; DNA; 14024 BP.
XX
AC   AAZ56386;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:56.
XX
KW   Flagellin; fliC; antigen; detection; ds.
XX
OS   Escherichia coli.
XX
PN   WO9961458-A1.
XX
PD   02-DEC-1999.
XX
PF   21-MAY-1999;   99WO-AU000385.
XX
PR   21-MAY-1998;   98AU-00003634.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 2000-072598/06.
XX
PT   Novel nucleic acid molecule useful for the detection of flagellated
PT   bacterial strains in food, feces, etc.
XX
PS   Claim 3; Page 229-233; 245pp; English.
XX
CC   AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all or
CC   part of an Escherichia coli flagellin protein except a protein expressed
CC   by E. coli H1, H7, H12 or H48 type strains. The present invention also
CC   describes a method of detecting the presence of E. coli of a particular H
CC   serotype in a sample, comprising specifically hybridising a nucleic acid,
CC   preferably at least a pair, derived from a flagellating gene, specific
CC   for a particular flagellin gene associated with the H serotype, to any
CC   E.coli in the sample which contain the gene, and detecting any hybridised
CC   molecules, identifying the presence of that serotype in the sample. (I)
CC   are useful for: (1) detecting the presence of E. coli of H serotype in a
CC   sample by hybridising at least one or a pair of (I) to any E. coli in the
CC   sample and detecting the hybridised nucleic acid molecules; and (2) for
CC   detecting the presence of both O and H-serotypes of E. coli by
CC   hybridising at least one or a pair of (I) to any E. coli present in the
CC   sample and detecting the hybridised nucleic acid molecules. (I) is
CC   particularly useful for detecting the combination of O and H antigen.
CC   Hybridised (I) when using at least one (I) is detected by southern blot
CC   analysis and, when using a pair of (I), is detected by polymerase chain
CC   reaction (PCR). AAZ56399 to AAZ56420 represent primers used in the
CC   exemplification of the present invention
XX
SQ   Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 1185;  DB 3;  Length 14024;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-205;
  Matches 1185;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 917

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 977

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 1037

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 1097

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 1157

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 1217

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 1277

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 1337

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 1397

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 1457

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 1517

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 1577

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1578 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 1637

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1638 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 1697

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1698 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 1757

Qy        901 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1758 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 1817

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1818 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1877

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1878 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1937

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1938 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1997
```

```
Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1998 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 2042
```

RESULT 3
ADC01214/c
ID ADC01214 standard; DNA; 14181 BP.
XX
AC ADC01214;
XX
DT 04-DEC-2003 (first entry)
XX
DE Enterohaemorragic E. coli 0157:H7-specific nucleic acid SEQ ID NO: 1258.
XX
KW ds; gene; enterohaemorragic; anti-bacterial.
XX
OS Escherichia coli; 0157:H7.
XX
PN JP2002355074-A.
XX
PD 10-DEC-2002.
XX
PF 24-JAN-2002; 2002JP-00015959.
XX
PR 24-JAN-2001; 2001JP-00112010.
XX
PA (UYTS-) UNIV TSUKUBA.
XX
DR WPI; 2003-451640/43.
XX
PT Enterohemorragic Escherichia coli O157:H7-specific nucleic acid molecule
PT and a polypeptide and its use, a polypeptide, a vector and a host cell.
XX
PS Claim 2; SEQ ID NO 1258; 2067pp; Japanese.
XX
CC The invention relates to a novel enterohaemorragic Escherichia coli
CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
CC has anti-bacterial activity. The polypeptide can be used in detection
CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
CC genome of Enterohaemorragic E coli O157:H7 was determined. The present
CC sequence represents an E. coli 0157:H7-specific nucleic acid of the
CC invention.
XX
SQ Sequence 14181 BP; 4450 A; 2856 C; 2281 G; 4594 T; 0 U; 0 Other;

Query Match 99.9%; Score 1183.4; DB 10; Length 14181;
Best Local Similarity 99.9%; Pred. No. 9e-205;
Matches 1184; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13138 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 13079

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13078 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 13019

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13018 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 12959

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12958 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 12899

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12898 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 12839

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12838 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 12779

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12778 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 12719

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12718 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 12659

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12658 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 12599

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12598 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 12539

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12538 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 12479

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      12478 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAAGTGTTCCATATGTTGTTTCT 12419

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12418 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 12359

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12358 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 12299

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12298 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 12239

Qy        901 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12238 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 12179

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12178 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 12119

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12118 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 12059

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12058 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 11999

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||||||||||||||||||||||||||||||||||||||||
Db      11998 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 11954
```

RESULT 4
ACD19108/c
ID ACD19108 standard; DNA; 14187 BP.
XX
AC ACD19108;
XX
DT 27-OCT-2003 (revised)
DT 21-AUG-2003 (first entry)
XX
DE E. coli 0157 unique DNA sequence OZID_121.
XX
KW OZID; ds; acute haemorrhagic colitis; haemolytic uraemic syndrome;
KW food poisoning.
XX
OS Escherichia coli; strain 0157:H7.

```
XX
PN   US2003023075-A1.
XX
PD   30-JAN-2003.
XX
PF   01-APR-2002; 2002US-00114170.
XX
PR   04-DEC-1998;   98US-0110955P.
PR   03-DEC-1999;   99US-00453702.
XX
PA   (BLAT/) BLATTNER F R.
PA   (BURL/) BURLAND V D.
PA   (PERN/) PERNA N T.
PA   (PLUN/) PLUNKETT G.
PA   (WELC/) WELCH R.
XX
PI   Blattner FR,  Burland VD,  Perna NT,  Plunkett G,  Welch R;
XX
DR   WPI; 2003-479497/45.
XX
PT   New DNA sequences from Escherichia coli strain O157:H7, useful for
PT   detecting E. coli O157:H7 in a sample, or in designing diagnostic probes
PT   which can be used to distinguish strain O157:H7 from strain K12 using
PT   molecular techniques.
XX
PS   Claim 16; SEQ ID NO 121; 33pp; English.
XX
CC   The invention relates to an isolated DNA molecule comprising an E. coli
CC   strain 0157:H7 sequence selected from a clostridial cytotoxin-like gene,
CC   a urease gene cluster, a RTX toxin-like gene cluster, a locus of
CC   enterocyte effacement and 2 genes from its associated lymphocytic phage
CC   933W (a putative serine/threonine kinase and a tail fibre gene). E. coli
CC   0157:H7 can cause food poisoning, specifically acute haemorrhagic colitis
CC   (which can develop into haemolytic uraemic syndrome). Also included are
CC   an isolated DNA molecule comprising a nucleotide sequence identical to at
CC   least 25 contiguous nucleotides contained in DNA sequences selected from
CC   ACD18988-ACD19242 (being 255 E.coli 0157 DNA sequences which are not
CC   found in E.coli K12), a recombinant DNA construction comprising the DNA
CC   above and a method for detecting E. coli 0157:H7 (ATCC 43895) in a sample
CC   (or distinguishing between 0157 and K12) using a probe derived from one
CC   of the 255 sequences. The DNA sequences are useful in detecting E. coli
CC   O157:H7 in a sample, for the early diagnosis of humans and livestock
CC   infected with O157:H7, and in designing diagnostic probes which can be
CC   used to distinguish strain O157:H7 from strain K12 using molecular
CC   techniques. The present sequence is one of the 255 E. coli 0157:H7 DNA
CC   sequence (termed OZID_1-OZID255). Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from the USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030023075 (Updated on 27-OCT-2003
CC   to standardise OS field)
XX
SQ   Sequence 14187 BP; 4451 A; 2859 C; 2282 G; 4595 T; 0 U; 0 Other;

  Query Match             99.9%;  Score 1183.4;  DB 9;  Length 14187;
  Best Local Similarity   99.9%;  Pred. No. 9e-205;
  Matches 1184;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13143 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 13084

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13083 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 13024

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13023 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 12964

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12963 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 12904

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12903 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 12844

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12843 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 12784

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12783 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 12724

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12723 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 12664

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12663 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 12604

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12603 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 12544

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12543 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 12484

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      12483 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAAGTGTTCCATATGTTGTTTCT 12424

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12423 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 12364

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12363 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 12304

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12303 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 12244

Qy        901 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12243 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 12184

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12183 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 12124

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12123 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 12064

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12063 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 12004

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||||||||||||||||||||||||||||||||||||||
Db      12003 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 11959


RESULT 5
ACA64845
ID   ACA64845 standard; DNA; 115218 BP.
XX
AC   ACA64845;
XX
DT   27-JUN-2003  (first entry)
XX
```

DE Human HNRNP GP43 DNA corresponding to AL034397.
XX
KW Human; chronic inflammatory joint disease; infection; tumour;
KW antiinflammatory; cytostatic; antiarthritic; antirheumatic;
KW immunosuppressive; gene therapy; etiological pathogenicity; ds.
XX
OS Homo sapiens.
XX
PN DE10127572-A1.
XX
PD 05-DEC-2002.
XX
PF 30-MAY-2001; 2001DE-01027572.
XX
PR 30-MAY-2001; 2001DE-01027572.
XX
PA (PATH-) PATHOARRAY GMBH.
XX
PI Haeupl T, Ungethuem U, Blaess S;
XX
DR WPI; 2003-240797/24.
XX
PT Reagents for diagnosis, study and therapy of chronic inflammatory joint
PT and other diseases, comprises any of many specified genes or derived
PT proteins.
XX
PS Claim 1; Page; 12pp; German.
XX
CC This invention describes a novel reagent for diagnosis, molecular
CC definition and therapy of chronic inflammatory joint diseases, and other
CC inflammatory disorders, infective or tumour diseases in humans. The
CC products of the invention have antiinflammatory, cytostatic,
CC antiarthritic, antirheumatic and immunosuppressive activity and can be
CC used for gene therapy. The reagent of the invention and any proteins and
CC antibodies derived from it, are used (i) for analysing tissue and blood
CC samples for medical diagnosis; (ii) for diagnosis and characterisation of
CC chronic joint diseases, on the basis of molecular characterisation, and
CC determining the etiological pathogenicity principle of as yet
CC uncharacterised inflammatory diseases, also monitoring progression and/or
CC treatment of disease, and optimisation of therapy and (iii) for
CC developing treatments for inflammatory diseases, particularly of joints,
CC infections and tumours. ACA64801-ACA64965 represent human polynucleotides
CC used in the method of the invention
XX
SQ Sequence 115218 BP; 35618 A; 23477 C; 22286 G; 33837 T; 0 U; 0 Other;

Query Match 6.3%; Score 75.2; DB 8; Length 115218;
Best Local Similarity 43.9%; Pred. No. 0.00027;
Matches 428; Conservative 0; Mismatches 538; Indels 10; Gaps 2;

```
Qy        74 TAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTC 133
             || || |   |||||| | | | | ||| |     | ||  |  |||||||  |  ||||
Db     16655 TATATAAATATATAATATAATATAATATTATGAATATATTATATTATATTATTATTAT 16714

Qy       134 AGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAA 193
             | | | | | |  | |||| | |||| |||  |||||| ||   | |  |    | ||
Db     16715 ATAATATATATTATATATTGTATTATATAATTATTATATATATTATATTATTATATAATT 16774

Qy       194 AATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAA 253
             |  ||||      |          ||    |  ||| |   | |   | || | ||   | | |
Db     16775 ATATTATATATTATATTATTATATGATATATATTATATTAATATTATTACATAATATA 16834

Qy       254 TTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGG 313
             |      | |  | |||| |  | |   |     |  ||||| | ||  |      |
Db     16835 TATTATATATTATATTATATTATTATATAATATATATTATATATTATATTATATTATTAT 16894

Qy       314 GGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTG 373
               || |  |   |||     |   | ||  | | || |||| ||||      ||||  |
Db     16895 ATAATATATATTATATATTATATTATTATATAATATATATTATATATTA----TATTATA 16950

Qy       374 GTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGA 433
             || ||||        |   ||  |||||  | ||| ||  |    | ||  |
Db     16951 TTATTATATAATATATATTATATATTATATTATTATTATATAATATATATTATATATT 17010

Qy       434 CCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGT 493
             || |||||  |||||  ||| |       |  |||| |  |  |  |    |||   | |
Db     17011 ATATTATATTATTATATATTATATATTATATATTATATTATTATTATATAATATATAT 17070

Qy       494 TGCTCATTCT------TCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTA 547
             |    ||| |      | | |||||| |  ||   | |||  | |   |  |||
Db     17071 TATATATTATATTATATTATTATATAATATTTTATATTATATTATATTATTATATAATAT 17130

Qy       548 ATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGT 607
             ||  || |  | | |  ||| ||  |||  | |||       | |     |       |
Db     17131 ATATTATATGATATATTATATTATTATATAATATATATTATATGATATATTATATTACAT 17190

Qy       608 TTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGA 667
             | ||| | |  |||     |  | |  | |     |||| |    |     |  | |
Db     17191 TATATATTATATTATATGATATATTATATGATATATTATATTATATGATATATTATATTA 17250

Qy       668 TAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCTA 727
             ||  | |||   ||  |     |   |     |  | |||||  | |     |   ||
Db     17251 TATATTATATAATATAATATAATATATCATATAATATATCATATAATATAATATATCATA 17310

Qy       728 TTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATAATAAATA 787
             | |   |  || |  |           |  | | | |    || | |   || ||| |||
Db     17311 TAATATATTATATAATAAAATATATAATATAATATATAATATAATATATTATTATATATA 17370

Qy       788 GAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAACAAGTAT 847
             |||| |   ||||   || |    |   | ||  | ||   |    | | | |
Db     17371 TAATATATTATATAACATATTATATATATAATAATATATAATATTATATAATATATTATA 17430

Qy       848 TTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTGTTAGTTT 907
             ||  | | |    | | | | | ||| |  | ||           ||    |  |
Db     17431 TTATATATAATATAATATAATATATTATATTATATTATATATTATTAATATATTATAA 17490

Qy       908 TTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTTTATCTTT 967
             || |    | | | | |    | |||||     |       ||||||    || ||| | |
Db     17491 TTATATAATATATATATTATATAATTATAATATATTAATAATATATCGTATTATATATAT 17550

Qy       968 TAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATATATCATTA 1027
             || |    || | | |       ||    |||   |   ||| ||  ||||   ||||
Db     17551 TATATTATATTATATATTATATTATTAAAATATTATCTATATTATAAATATATTATATTA 17610

Qy      1028 TGATTTTATTGCAATA 1043
             |  |||||||  | ||
Db     17611 TATTTTTATTATATTA 17626
```

RESULT 6
ACA64845/c
ID ACA64845 standard; DNA; 115218 BP.
XX
AC ACA64845;
XX
DT 27-JUN-2003 (first entry)
XX
DE Human HNRNP GP43 DNA corresponding to AL034397.
XX
KW Human; chronic inflammatory joint disease; infection; tumour;
KW antiinflammatory; cytostatic; antiarthritic; antirheumatic;
KW immunosuppressive; gene therapy; etiological pathogenicity; ds.
XX
OS Homo sapiens.
XX
PN DE10127572-A1.
XX
PD 05-DEC-2002.
XX
PF 30-MAY-2001; 2001DE-01027572.
XX
PR 30-MAY-2001; 2001DE-01027572.
XX
PA (PATH-) PATHOARRAY GMBH.
XX
PI Haeupl T, Ungethuem U, Blaess S;

XX
DR WPI; 2003-240797/24.
XX
PT Reagents for diagnosis, study and therapy of chronic inflammatory joint
PT and other diseases, comprises any of many specified genes or derived
PT proteins.
XX
PS Claim 1; Page; 12pp; German.
XX
CC This invention describes a novel reagent for diagnosis, molecular
CC definition and therapy of chronic inflammatory joint diseases, and other
CC inflammatory disorders, infective or tumour diseases in humans. The
CC products of the invention have antiinflammatory, cytostatic,
CC antiarthritic, antirheumatic and immunosuppressive activity and can be
CC used for gene therapy. The reagent of the invention and any proteins and
CC antibodies derived from it, are used (i) for analysing tissue and blood
CC samples for medical diagnosis; (ii) for diagnosis and characterisation of
CC chronic joint diseases, on the basis of molecular characterisation, and
CC determining the etiological pathogenicity principle of as yet
CC uncharacterised inflammatory diseases, also monitoring progression and/or
CC treatment of disease, and optimisation of therapy and (iii) for
CC developing treatments for inflammatory diseases, particularly of joints,
CC infections and tumours. ACA64801-ACA64965 represent human polynucleotides
CC used in the method of the invention
XX
SQ Sequence 115218 BP; 35618 A; 23477 C; 22286 G; 33837 T; 0 U; 0 Other;

Query Match 6.3%; Score 74.8; DB 8; Length 115218;
Best Local Similarity 43.6%; Pred. No. 0.00032;
Matches 428; Conservative 0; Mismatches 552; Indels 2; Gaps 2;

```
Qy        69 TATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTAT 128
             ||| ||| ||         | | | | |  ||  ||  || ||  || ||| |||
Db     17607 TATAATATATTTATAATATAGATAATATTTTAATAATATAATATATAATATAATATAATA 17548

Qy       129 ATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCA 188
               |  |       | |  | ||| ||| ||||      || ||| ||  |  | |  |  |
Db     17547 TATATAATACGATATATTATTAATATATTATAATTATATATAATATATATTATTATATAATTA 17488

Qy       189 GAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTAT 248
              || | ||| ||  |  |        || | |  ||| |   ||   || |    | |||
Db     17487 TAATATATTAATAATATATAATATAATATAATATATATTATATTATATTATATATAATAT 17428

Qy       249 AGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCC 308
             |  |  |  |    || | |   || ||| | |    ||   |    |     || ||
Db     17427 AATATATTATATAATATTATATATTATTATATATATAATATGTTATATAATATATTATAT 17368

Qy       309 AGGGGGGGCTAAACT-TAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGT 367
             |       || | | ||  |||  | |  ||      ||   ||   ||||||||  |
Db     17367 ATATAATAATATATTATATTATATATTATATTATATATTTTATTATATAATATATTATAT 17308

Qy       368 ATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAAC 427
               | |  ||   ||  ||   |  ||    |||  | ||| ||       |||  | |
Db     17307 GATATATTATATTATATGATATATTATATGATATATTATATTATATTATATAATATATAA 17248

Qy       428 AGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAAT 487
                 |||||||      || |  ||   | |   | || | |      | | |
Db     17247 TATAATATATCATATAATATAATATATCATATAATATATCATATAATATAATATATAATG 17188

Qy       488 TGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTA 547
             |  | |     ||  |  ||||||||||   ||  |||||  | |   |    |    ||
Db     17187 TAATATAATATATCATATAATATATATTATATAATAATATAATATATCATATAATATATA 17128

Qy       548 ATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGT 607
               |    ||  | | |  ||||||    || || |||     |      | ||   |
Db     17127 TTATATAATAATATAATATAATATAAAATATTATATAATAATATAATATAATATATAATA 17068

Qy       608 TTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGA 667
             | |||    |   ||  | || |       | || ||     ||||||||||||  | |  |
Db     17067 TATATTATATAATAATATAATATAATATATAATATATAATATATAATAATATAAT-ATAA 17009

Qy       668 TAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCTA 727
             ||  |||||  |     |||     |    |   | |   |||| |  | | | |   ||
```

```
Db     17008 TATATAATATATATTATATAATAATATAATATAATATATAATATATATTATATAATAATA 16949

Qy       728 TTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATAATAAATA 787
             | |   |  || | |     |     | | | | | ||   || | |   | ||   ||||
Db     16948 TAATATAATATATAATATATATTATATAATAATATAATATATAATATATATTATATAATA 16889

Qy       788 GAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAACAAGTAT 847
               |||     |   |       |     |||  ||     |     || ||||   |   |
Db     16888 ATATAATATAATATATAATATATATTATATAATAATATAATATAATATATAATATATATT 16829

Qy       848 TTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTGTTAGTTT 907
               || |   ||   |  | | ||  | |||    ||  |       |     | | | | | |
Db     16828 ATGTAATAATATTAATATATAATATATATCATATAATAATATAATATATAATATAATTAT 16769

Qy       908 TTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTTTATCTTT 967
               |  | | |  | | | || | |||||     |     |||||       || |||  |
Db     16768 ATAATAATATAATATATATAATAATTATATAATACAATATATAATATATATTATATAATA 16709

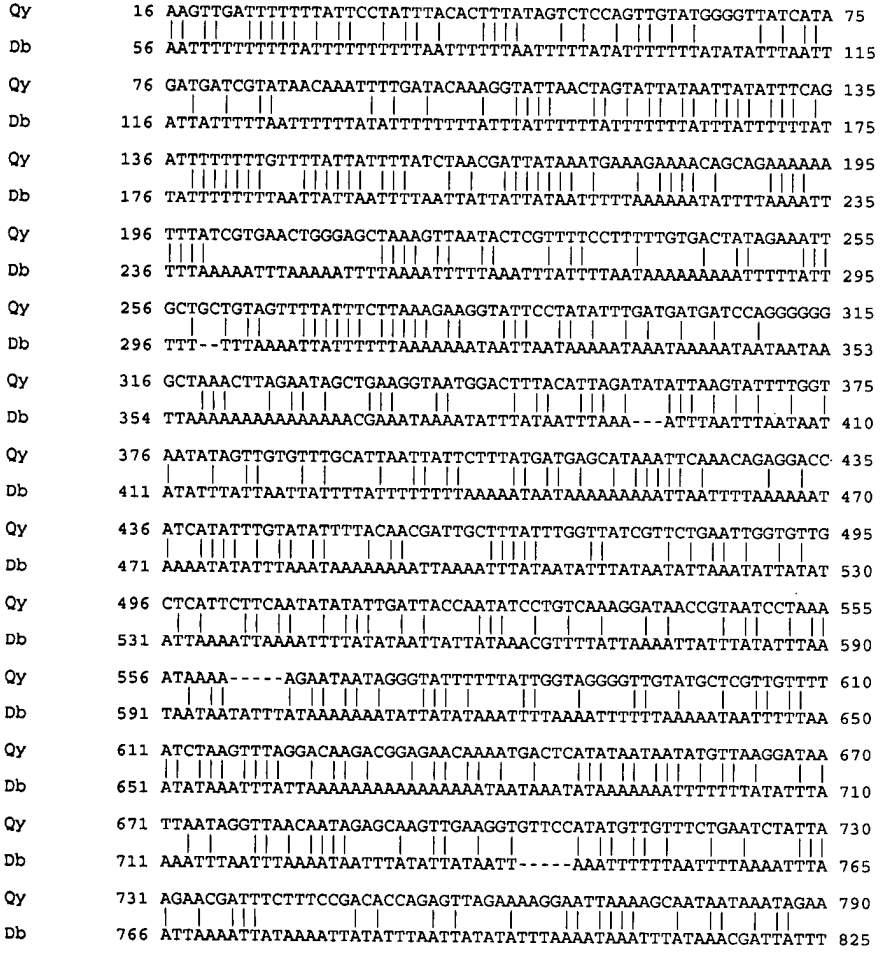
Qy       968 TAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATATATCATTA 1027
               || | |||  || | |  ||||      || |||      | |||| |||||    |
Db     16708 ATAATATAATATAATATATTCATAATATATTATATTATATTATATATTTATATACTTATT 16649

Qy      1028 TGATTTTATTGCAATACTTAGT 1049
               ||| |||| |     ||| |
Db     16648 ATATTATATTACTTCCTTTATT 16627
```

RESULT 7
ABZ10100
ID ABZ10100 standard; DNA; 8056 BP.
XX
AC ABZ10100;
XX
DT 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #240.
XX
KW Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 240; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118

CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;

Query Match 6.0%; Score 71; DB 8; Length 8056;
Best Local Similarity 45.0%; Pred. No. 0.0013;
Matches 438; Conservative 0; Mismatches 520; Indels 15; Gaps 4;





RESULT 8
ABZ10246
ID ABZ10246 standard; DNA; 8056 BP.
XX
AC ABZ10246;
XX
DT 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #386.
XX
KW Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 386; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA

```
CC   sequences. The nucleotide sequences from the present invention can also
CC   be used for detecting a predisposition to, differentiation between
CC   subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC   haematopoietic cell proliferative disorders. The present method enables a
CC   highly specific classification of haematopoietic cell proliferative
CC   disorders allowing for improved and informed treatment of patients
XX
SQ   Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;

  Query Match             5.9%;  Score 70;  DB 8;  Length 8056;
  Best Local Similarity   44.5%;  Pred. No. 0.002;
  Matches  502;  Conservative    0;  Mismatches  610;  Indels   16;  Gaps    5;

Qy         16 AAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATA 75
Db         56 AATTTTTTTTTTATTTTTTTTTTAATTTTTTAATTTTTATATTTTTTTATATATTTAATT 115

Qy         76 GATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAG 135
Db        116 ATTATTTTTAATTTTTTATATTTTTTTTATTTATTTTTTATTTTTTTATTTATTTTTTAT 175

Qy        136 ATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAA 195
Db        176 TATTTTTTTTAATTATTAATTTTAATTATTATTATAATTTTTAAAAAATATTTTAAAATT 235

Qy        196 TTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATT 255
Db        236 TTTAAAAATTTAAAAATTTTAAAATTTTTAAATTTATTTTAATAAAAAAAAATTTTTATT 295

Qy        256 GCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGG 315
Db        296 TTT--TTTAAAATTATTTTTTAAAAAAATAATTAATAAAAATAAATAAAAATAATAATAA 353

Qy        316 GCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGT 375
Db        354 TTAAAAAAAAAAAAAAAATGAAATAAAATATTTATAATTTAAA---ATTTAATTTAATAAT 410

Qy        376 AATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACC 435
Db        411 ATATTTATTAATTATTTTATTTTTTTTAAAAATAATAAAAAAAAATTAATTTTAAAAAAT 470

Qy        436 ATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTG 495
Db        471 AAAATATATTTAAATAAAAAAAATTAAAATTTATAATATTTATAATATTAAATATTATAT 530

Qy        496 CTCATTCTTCAATATATATTGATTACCAATA-----TCCTGTCAAAGGATAACCGTAATC 550
Db        531 ATTAAAATTAAAATTTTATATAATTATTATAAATGTTTTATTAAAATTATTTATATTTAA 590

Qy        551 CTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTT 610
Db        591 TAATAATATTTATAAAAAAATATTATATAAATTTTAAAATTTTTTAAAAATAATTTTTAA 650

Qy        611 ATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAA 670
Db        651 ATATAAATTTATTAAAAAAAAAAAAAAAAATAATAAATATAAAAAAATTTTTTTATATTTA 710

Qy        671 TTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCTATTA 730
Db        711 AAATTTAATTTAAAATAATTTATATTATAATT-----AAATTTTTTAATTTTAAAATTTA 765

Qy        731 AGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATAATAAATAGAA 790
Db        766 ATTAAAATTATAAAATTATATTTAATTATATATTTAAAATAAATTTATAAATGATTATTT 825

Qy        791 TACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAACAAGTATTTG 850
Db        826 AAAATATTTTTATTAAAAAAAATTGTTTTTTTTATTTTTTTTTATGTATAAAAATTTTGT 885

Qy        851 GAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTGTTAGTTTTTT 910
Db        886 AATTATTTAAATTTATATTTTAATTATTTTGTTGTTTTTGTAATAAATTTTTTTTAAAA 945

Qy        911 TTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATA-TATACCTTTTTATCTTTTA 969
Db        946 ATAAATTTTTATTTTTTTATTTAAATGAATAAAATATATTATTATAATTTTTTTAAAAATA 1005

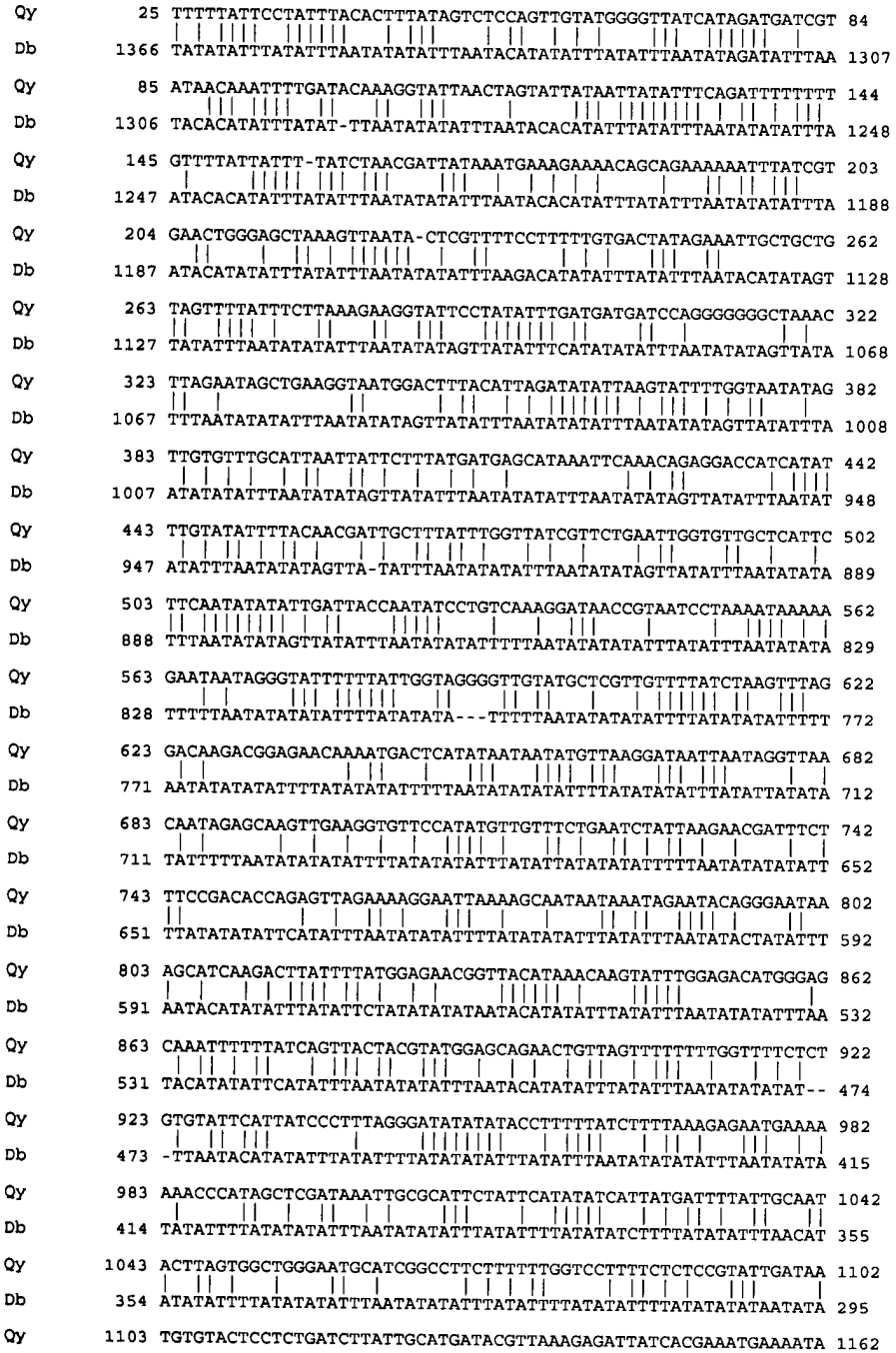
Qy        970 AAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATATATCATTATG 1029
Db       1006 ATTATTTTAAAAATTTTTTTATTATTTTTTAATTTTGAAATTTAATTATGTAAAATAAAA 1065

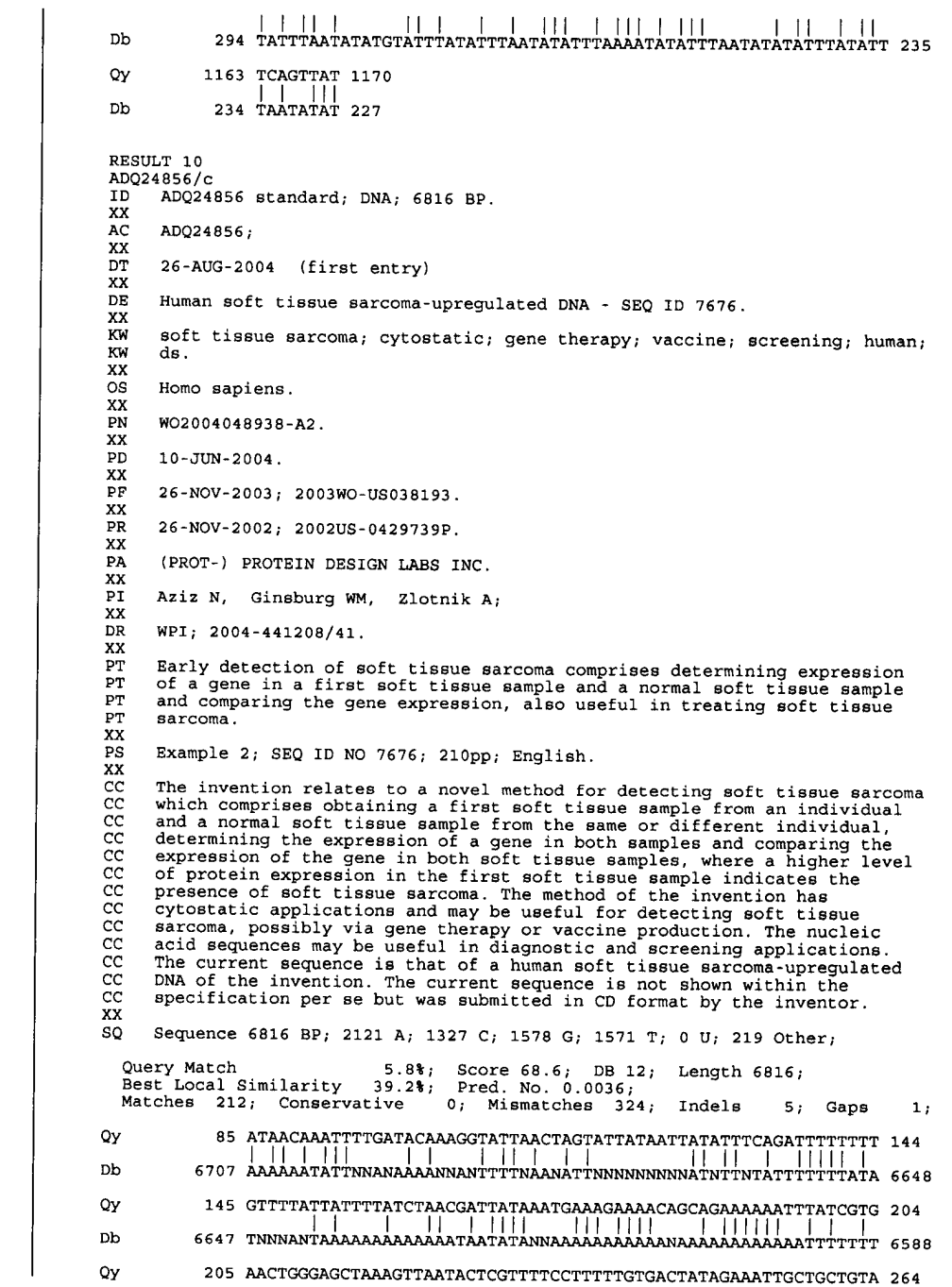
Qy       1030 ATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGTCCTTTTCTC 1089
Db       1066 AAATTTTGTTTTAATTTGTTTTTTTTTTAATGTTTTAAAATGTTATATATTTGTTTTTTA 1125

Qy       1090 TCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTA 1137
Db       1126 ATATTTTTTTTGATATTTGAAAATTTAAAATTAATATATTTTTGATTA 1173


RESULT 9
ADZ70999/c
ID   ADZ70999 standard; DNA; 1554 BP.
XX
AC   ADZ70999;
XX
DT   14-JUL-2005  (first entry)
XX
DE   Human chromosome 2 contig DNA SEQ ID NO 123.
XX
KW   ds; matrix attachment region; MAR; protein production.
XX
OS   Homo sapiens.
XX
PN   WO2005040377-A2.
XX
PD   06-MAY-2005.
XX
PF   22-OCT-2004; 2004WO-EP011974.
XX
PR   24-OCT-2003; 2003US-0513574P.
PR   06-FEB-2004; 2004EP-00002722.
XX
PA   (SELE-) SELEXIS SA.
XX
PI   Mermod N,  Girod PA,  Bucher P,  Nguyen D,  Calabrese D,  Saugy D;
PI   Puttini S;
XX
DR   WPI; 2005-333507/34.
XX
PT   New purified and isolated DNA sequence having protein production
PT   increasing activity comprises a bent DNA element and a binding site for a
PT   DNA binding protein, useful for increasing protein production activity in
PT   eukaryotic host cell.
XX
PS   Disclosure; SEQ ID NO 123; 282pp; English.
XX
CC   The invention relates to a purified and isolated DNA sequence having
CC   protein production increasing activity comprising at least one bent DNA
CC   element, and at least one binding site for a DNA binding protein. The
CC   purified and isolated DNA sequence comprising a first and a second
CC   isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC   MAR nucleotide sequence selected from a purified and isolated DNA
CC   sequence above, a purified and isolated MAR DNA above, a purified and
CC   isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC   sequence complementary to it, its molecular chimera, or its combinations
CC   and variants, is useful for increasing protein production activity in a
CC   eukaryotic host cell. The present sequence represents a human chromosome
CC   2 contig DNA.
XX
SQ   Sequence 1554 BP; 865 A; 14 C; 27 G; 648 T; 0 U; 0 Other;

  Query Match             5.9%;  Score 69.6;  DB 14;  Length 1554;
  Best Local Similarity   44.8%;  Pred. No. 0.0022;
  Matches  514;  Conservative    0;  Mismatches  624;  Indels   10;  Gaps    6;
```





RESULT 10
ADQ24856/c
ID ADQ24856 standard; DNA; 6816 BP.
XX
AC ADQ24856;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7676.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 7676; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 6816 BP; 2121 A; 1327 C; 1578 G; 1571 T; 0 U; 219 Other;

Query Match 5.8%; Score 68.6; DB 12; Length 6816;
Best Local Similarity 39.2%; Pred. No. 0.0036;
Matches 212; Conservative 0; Mismatches 324; Indels 5; Gaps 1;

```
Db       6587 TATAAATTTTTTTTAAAAAAAAAANNTTTTNTTTTAAAAAATATTTAAATANNNTTTTTA 6528

Qy        265 GTTTTATTTCTTAAAGAAGGTATTCCTATATTTG-----ATGATGATCCAGGGGGGGCTA 319

Db       6527 TTTTTAATTTNTTTATTTTAAAAATAAAAATATNTAAAAAAANTNATTTTTTTTAAANAA 6468

Qy        320 AACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATA 379

Db       6467 AANTNAAAATATTTTTTTTTTATTTTTTTNCCNNAAAAAATAATAATAANTNANNTAAAA 6408

Qy        380 TAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCA 439

Db       6407 AATTTTTTTTTTAAATAAANTNTTTTATAAAAAATATTTTTTTTAAAAAAAAAAANTAA 6348

Qy        440 TATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCA 499

Db       6347 AATTTNNTNNTTTTTTTATTNNNNTNTTATTTATTTTTTTNTNTTTNNNNNNNNNTTT 6288

Qy        500 TTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATAA 559

Db       6287 TTTTTTNNNNTNTTTTTTTTTTTTTTNTTNNNNNTTTNTNTTTNNNNTNNNNNNTTTNNN 6228

Qy        560 AAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGTT 619

Db       6227 NNNNNTNNNNNNNTTTTTNTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTT 6168

Qy        620 T 620

Db       6167 T 6167
```

RESULT 11
ABL32101
ID ABL32101 standard; DNA; 5728 BP.
XX
AC ABL32101;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 74.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 74; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 5728 BP; 1931 A; 49 C; 885 G; 2863 T; 0 U; 0 Other;

Query Match 5.8%; Score 68.4; DB 6; Length 5728;
Best Local Similarity 48.8%; Pred. No. 0.0038;
Matches 244; Conservative 0; Mismatches 251; Indels 5; Gaps 2;

```
Qy         19 TTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATAGAT 78

Db       3956 TTTATGTTATTTTTTATTTTGATAGGTTATATTTTTGATTTTTTTTTTTTTAAAGATAAA 4015

Qy         79 GATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTA---TTATAATTATATTTCAG 135

Db       4016 GGGAAATATATTTATGGAAATGTAATAATATTTAATAATATTTTTAATATTTTATAAGAG 4075

Qy        136 ATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAA 195

Db       4076 ATTTTTTATAAATAATTATTTTATTTAGAGTTATTTTAAAATATAAGAAATTTGGATTTG 4135

Qy        196 TTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATT 255

Db       4136 TTTGGAGAGAAAATTGGGGATAGGAAAATTTTATTATTTTTTTTATTATAAAATAAATATA 4195

Qy        256 GCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGG 315

Db       4196 TGTAGATTTTTTTTATTATGAAGAGTTTGTATTATTTTTTATAA--AATATGTAAGAGTG 4253

Qy        316 GCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGT 375

Db       4254 GAAAAGTGAAAGATTGTTAAGTATTAAATTAATTAATTGAAATATAATTAGTAATTTAAT 4313

Qy        376 AATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACC 435

Db       4314 TTTTTATTTATTTATGGAGATATATTTTTTTTAAAAAAAGTTAGAATGAAAGTGTATAAA 4373

Qy        436 ATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTG 495

Db       4374 TTTATATAAGTATTTTTAATGTAAATAAAAATGTATAAATATATAGGTATTTTAATGTAA 4433

Qy        496 CTCATTCTTCAATATATATT 515

Db       4434 ATAAAAAATTGGAATATAAT 4453
```

RESULT 12
AEB32373/c
ID AEB32373 standard; DNA; 38678 BP.
XX
AC AEB32373;
XX
DT 08-SEP-2005 (first entry)
XX
DE Human genomic DNA #14.
XX
KW SNP detection; diagnosis; non-insulin dependent diabetes; obesity;
KW antidiabetic; anorectic; endocrine disease; gastrointestinal disease;
KW metabolic disorder; nutritional disorder; gene; ds.
XX
OS Homo sapiens.
XX
PN US2005147987-A1.
XX
PD 07-JUL-2005.
XX

PF 19-JUL-2004; 2004US-00893315.
XX
PR 08-SEP-2000; 2000US-0231397P.
PR 10-SEP-2001; 2001US-00948947.
XX
PA (APPL-) APPLERA CORP NY.
XX
PI Venter JC, Zhang JN, Liu X, Rowe W, Cravchik A, Kalush F;
PI Naik A, Subramanian G, Woodage T;
XX
DR WPI; 2005-511776/52.
XX
PT New detection reagent capable of detecting 1, 100, 500, 1000 or 5000 or
PT more single nucleic acid polymorphisms, useful in identifying an
PT individual having or at risk of developing type II diabetes or obesity.
XX
PS Disclosure; SEQ ID NO 136; 31pp; English.
XX
CC The invention relates to a detection reagent capable of detecting one or
CC more single nucleic acid polymorphisms. The invention also relates to
CC determining whether a trait is linked to one of the human chromosomes or
CC its sub-region, a computer readable medium having stored in it the SNP
CC relational information given in the specification, an isolated nucleic
CC acid molecule for detecting at least one SNP given in the specification
CC comprising at least about 12 contiguous nucleotides, genotyping at least
CC one SNP position given in the specification in a sample, identifying an
CC individual having or at risk of developing a disorder and a kit
CC comprising at least one container containing the detection reagent.
CC Determining whether a trait is linked to one of the human chromosomes or
CC its sub-region comprises determining whether the trait is linked to one
CC or more SNPs using the detection reagents. Genotyping at least one SNP
CC position given in the specification in a sample comprises contacting the
CC sample with a detection reagent that differentiates between alternative
CC alleles at at least one SNP position given in the specification, and
CC determining which allele is present at the at least one SNP position.
CC Identifying an individual having or at risk of developing a disorder
CC comprises genotyping at least one SNP given in the specification in a
CC nucleic acid sample from the individual. The disorder is type II diabetes
CC (non-insulin dependent diabetes) or obesity. The detection reagent is
CC useful in identifying an individual having or at risk of developing a
CC disorder, particularly type II diabetes or obesity. This sequence
CC represents human genomic DNA used in the scope of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 38678 BP; 9340 A; 9040 C; 9074 G; 10537 T; 0 U; 687 Other;

Query Match 5.7%; Score 68; DB 14; Length 38678;
Best Local Similarity 44.4%; Pred. No. 0.005;
Matches 319; Conservative 0; Mismatches 390; Indels 9; Gaps 2;

```
Qy      109 TTAACTAGTATTATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATT 168
            | |  ||  ||     ||||||| |     | ||  | | | ||   | | | ||   | |
Db    20575 TAATATATAATATATATTATATATTTATATATTAATATATAATATATATTTATATATAAT 20516

Qy      169 ATAAATGAAAGAAAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTC 228
            ||| || | | |           | | ||||||     | |   |  || | |||    |
Db    20515 ATATATTATATATTTATATTACATATATTTATATGTTAATATTACATATATTTATATATT 20456

Qy      229 GTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATT 288
              | |  ||| | |   |||| |   |      | ||  ||   ||  || | |  ||||
Db    20455 TATATATTTTATATATTTATATATTATATATTTATATATTATATTTATATATTATATATT 20396

Qy      289 CCTATATTTGATGATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTT 348
              |||  |  ||    ||          |   || | |    |  || | |       |
Db    20395 TATATTATATATTTATATTATATATTTATATATTATATATTTATATTATATATTTATATA 20336

Qy      349 TACATTAGATATATTAAGTATTTTGGTA----ATATAGTTGTGTTTGCATTAATTATTCT 404
            |    |  ||||||||  |||||   ||    |||||  || | | |    ||
Db    20335 TTGGATTTATATATTATATATTTATATACTATATATATTTATATATATTATANNNNNNNN 20276

Qy      405 TTATGATGAGCATAAATTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGC 464
```

```
                  || ||  |   | |       |||||| |||||| || |    |
Db    20275 NNNNNNNNNNNNTATATATATTTATATATATTATATATTTATATATTATATATATTTATA 20216

Qy      465 TTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAA 524
            | ||||  |  |  |    ||  | ||  |   | ||  |||| |||  ||||   |
Db    20215 TATATTATATATTTATATATTATATATATTTATATATATTATATATTTATATATTATATA 20156

Qy      525 TATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATT 584
            |||       | |  |||       ||  | | ||| |   |||  |    ||| || |||
Db    20155 TATTTATATATAATATATATTATATATTTATATATTATATATATTTATATATATTATATA 20096

Qy      585 GGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGA 644
              ||    || |||      |   | |||   |  ||   | |  |   | |   | ||
Db    20095 TTTATATATTATATATATTTATATATATTATATATTTATATATTATATATATTTATATAT 20036

Qy      645 CTCATATAATAATATGTTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGT 704
              | ||||| | |||| |||    ||| |||     |  | | || |   | ||  |  | |
Db    20035 ATTATATATTTATATATTATATATATTTA-----TATATATTATATATATTTATATATAT 19981

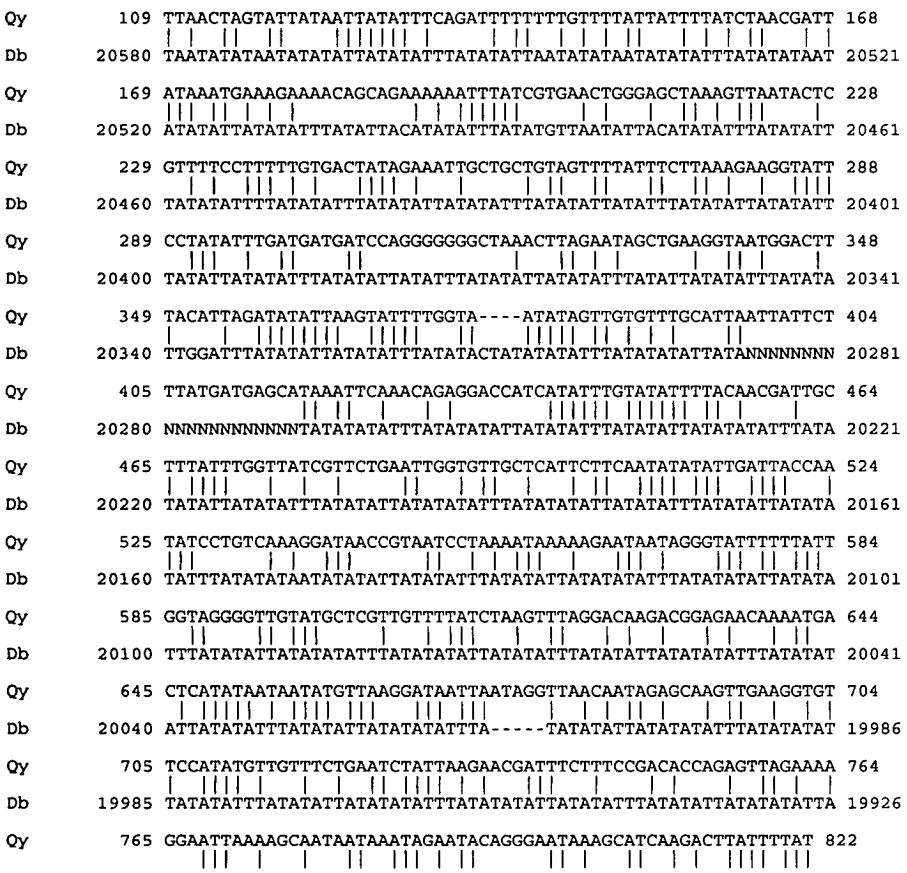
Qy      705 TCCATATGTTGTTTCTGAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAA 764
            |  |||| |   |  | || |||| | |   |||   |     | |     |  || |  |
Db    19980 TATATATTTATATATTATATATATTTATATATATTATATATTTATATATTATATATATTA 19921

Qy      765 GGAATTAAAAGCAATAATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTAT 822
              |||  |   |   ||  ||| | ||       || |   ||  | |  |||| |||
Db    19920 TATATTTTATCTATATATTTATATATTAATATATATTATATATTTATATATATTATAT 19863
```

RESULT 13
AEB32391/c
ID AEB32391 standard; DNA; 38684 BP.
XX
AC AEB32391;
XX
DT 08-SEP-2005 (first entry)
XX
DE Human genomic DNA #32.
XX
KW SNP detection; diagnosis; non-insulin dependent diabetes; obesity;
KW antidiabetic; anorectic; endocrine disease; gastrointestinal disease;
KW metabolic disorder; nutritional disorder; gene; ds.
XX
OS Homo sapiens.
XX
PN US2005147987-A1.
XX
PD 07-JUL-2005.
XX
PF 19-JUL-2004; 2004US-00893315.
XX
PR 08-SEP-2000; 2000US-0231397P.
PR 10-SEP-2001; 2001US-00948947.
XX
PA (APPL-) APPLERA CORP NY.
XX
PI Venter JC, Zhang JN, Liu X, Rowe W, Cravchik A, Kalush F;
PI Naik A, Subramanian G, Woodage T;
XX
DR WPI; 2005-511776/52.
XX
PT New detection reagent capable of detecting 1, 100, 500, 1000 or 5000 or
PT more single nucleic acid polymorphisms, useful in identifying an
PT individual having or at risk of developing type II diabetes or obesity.
XX
PS Disclosure; SEQ ID NO 154; 31pp; English.
XX
CC The invention relates to a detection reagent capable of detecting one or
CC more single nucleic acid polymorphisms. The invention also relates to
CC determining whether a trait is linked to one of the human chromosomes or
CC its sub-region, a computer readable medium having stored in it the SNP
CC relational information given in the specification, an isolated nucleic
CC acid molecule for detecting at least one SNP given in the specification
CC comprising at least about 12 contiguous nucleotides, genotyping at least

CC one SNP position given in the specification in a sample, identifying an
CC individual having or at risk of developing a disorder and a kit
CC comprising at least one container containing the detection reagent.
CC Determining whether a trait is linked to one of the human chromosomes or
CC its sub-region comprises determining whether the trait is linked to one
CC or more SNPs using the detection reagents. Genotyping at least one SNP
CC position given in the specification in a sample comprises contacting the
CC sample with a detection reagent that differentiates between alternative
CC alleles at at least one SNP position given in the specification, and
CC determining which allele is present at the at least one SNP position.
CC Identifying an individual having or at risk of developing a disorder
CC comprises genotyping at least one SNP given in the specification in a
CC nucleic acid sample from the individual. The disorder is type II diabetes
CC (non-insulin dependent diabetes) or obesity. The detection reagent is
CC useful in identifying an individual having or at risk of developing a
CC disorder, particularly type II diabetes or obesity. This sequence
CC represents human genomic DNA used in the scope of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 38684 BP; 9340 A; 9042 C; 9075 G; 10540 T; 0 U; 687 Other;

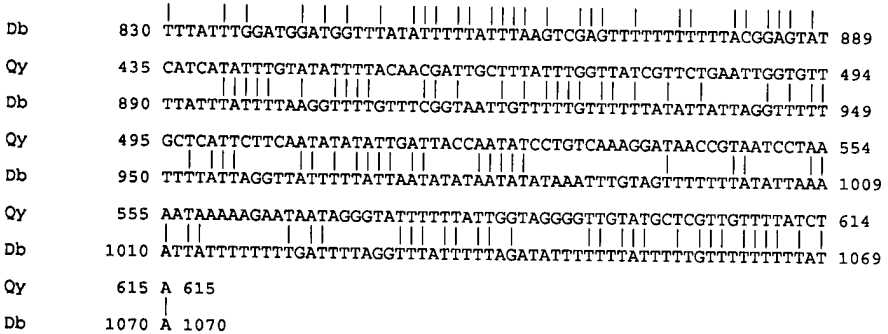
Query Match 5.7%; Score 68; DB 14; Length 38684;
Best Local Similarity 44.4%; Pred. No. 0.005;
Matches 319; Conservative 0; Mismatches 390; Indels 9; Gaps 2;



Db 19925 TATATTTTATCTATATATTTATATATTAATATATATTATATATTTATATATATTATAT 19868

RESULT 14
ABL34585
ID ABL34585 standard; DNA; 14551 BP.
XX
AC ABL34585;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human metastasis associated gene SEQ ID NO: 138.
XX
KW Metastasis associated gene; cytostatic; gene therapy; cancer;
KW cytosine methylation; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200177376-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP003970.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-010922/01.
XX
PT New nucleic acid derived from chemically treated metastasis genes, useful
PT for diagnosis of cancers by analysis of cytosine methylation, also for
PT treatment.
XX
PS Claim 1; SEQ ID NO 138; 23pp + Sequence Listing; English.
XX
CC The present invention provides a number of human metastasis associated
CC genes which are modified by cytosine methylation. The sequences can be
CC used in the diagnosis and treatment of cancer. The present sequence is
CC one of the genes of the invention
XX
SQ Sequence 14551 BP; 4132 A; 107 C; 2863 G; 7449 T; 0 U; 0 Other;

Query Match 5.7%; Score 67.8; DB 6; Length 14551;
Best Local Similarity 46.6%; Pred. No. 0.0052;
Matches 252; Conservative 0; Mismatches 287; Indels 2; Gaps 1;



```
Db        830 TTTATTTGGATGGATGGTTTATATTTTTATTTAAGTCGAGTTTTTTTTTTTACGGAGTAT 889

Qy        435 CATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTT 494

Db        890 TTATTTATTTTAAGGTTTTGTTTCGGTAATTGTTTTTGTTTTTTATATTATTAGGTTTTT 949

Qy        495 GCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAA 554

Db        950 TTTTATTAGGTTATTTTTATTAATATATAATATATAAATTTGTAGTTTTTTTATATTAAA 1009

Qy        555 AATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCT 614

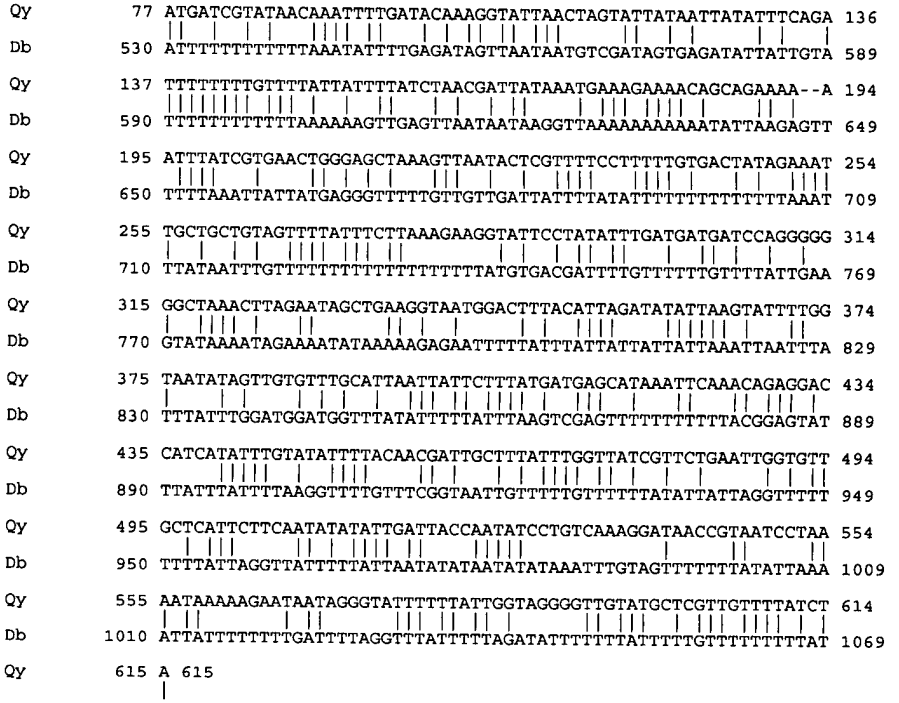
Db       1010 ATTATTTTTTTTGATTTTAGGTTTATTTTTAGATATTTTTTTATTTTTGTTTTTTTTTAT 1069

Qy        615 A 615

Db       1070 A 1070
```

RESULT 15
ADS99846
ID ADS99846 standard; DNA; 14551 BP.
XX
AC ADS99846;
XX
DT 02-DEC-2004 (first entry)
XX
DE Complement of bisulphite treated metastasis-associated human gene #69.
XX
KW Human; ds; gene; Bisulphite; metastasis; cancer; cytostatic;
KW DNA methylation; matrix-assisted laser desorption/ionisation; MALDI;
KW electrospray; mass spectrometry; CpG dinucleotide; solid tumour.
XX
OS Homo sapiens.
XX
PN US2003148327-A1.
XX
PD 07-AUG-2003.
XX
PF 21-JAN-2003; 2003US-00240485.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
PR 06-APR-2001; 2001WO-EP003970.
XX
PA (OLEK/) OLEK A.
PA (PIEP/) PIEPENBROCK C.
PA (BERL/) BERLIN K.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-010922/01.
XX
PT New nucleic acid derived from chemically treated metastasis genes, useful
PT for diagnosis of cancers by analysis of cytosine methylation, also for
PT treatment.
XX
PS Claim 1; SEQ ID NO 138; 9pp; English.
XX
CC The invention relates to a nucleic acid comprising at least 18 bases from
CC a segment of the chemically pretreated DNA of genes associated with
CC metastasis, i.e. any of ADS99709-ADS99906 human genomic sequences or any
CC of the 19 sequences appearing as ADS99911-ADS99929. SEQ ID 2,4,6 etc are
CC the complements of SEQ ID 1,3,5, etc. Also included are an oligomer
CC (particularly an oligonucleotide or peptide nucleic acid) comprising at
CC least one base sequence of at least 9 bases which hybridises to (or is
CC identical with) the sequences referred to above, producing an array of
CC the oligomers on a carrier, obtaining genetic and/or epigenetic
CC parameters for diagnosis and/or therapy of diseases (or predisposition to
CC them) by analysis of cytosine methylation and a kit comprising a
CC bisulphite (disulphite or hydrogen sulphite) and the oligomers. In the
CC method of above 5-unmethylated cytosines in a genomic DNA sample are
CC converted chemically to uracil, or another base with hybridisation
CC properties different from those of cytosine, then fragments of the
CC treated DNA amplified (particularly by polymerase chain reaction) using
CC the oligomers and a polymerase (preferably heat stable) to produce
CC labelled amplicons. These are tested for hybridisation to an array of
CC oligomers and any hybridisation detected. The amplicons are labelled with
CC fluorescent or radioactive markers, or with a detachable mass marker to
CC allow their detection by mass spectrometry, specifically using the matrix
CC -assisted laser desorption/ionisation (MALDI) or electrospray techniques.
CC To improve detection in the mass spectrometer, fragments formed in the
CC instrument have only a single net charge (positive or negative). The
CC genomic DNA is from e.g. a cell line, biopsy sample, blood, or paraffin-
CC embedded tissue sample. Oligonucleotides or peptide-nucleic acids that
CC are complementary to (or identical with) parts of the nuclei acids listed
CC above may be used as primers for amplification of the nucleic acids or
CC their complements, and for determining cytosine methylation status and/or
CC single nucleotide polymorphisms in metastasis-related genes. They can be
CC used for analysis of diseases associated with methylation of CpG
CC dinucleotides and to determine (epi)genetic parameters for diagnosis
CC and/or therapy of disease (or predisposition). The genomic DNA sequences
CC are useful for diagnosis and therapy of solid tumours and cancer. The
CC present sequence is the complementary sequence to a bisulphite treated
CC human gene associated with metastasis. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030148327.
XX
SQ Sequence 14551 BP; 4132 A; 107 C; 2863 G; 7449 T; 0 U; 0 Other;

Query Match 5.7%; Score 67.8; DB 7; Length 14551;
Best Local Similarity 46.6%; Pred. No. 0.0052;
Matches 252; Conservative 0; Mismatches 287; Indels 2; Gaps 1;



```
Qy         77 ATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGA 136

Db        530 ATTTTTTTTTTTTAAATATTTTGAGATAGTTAATAATGTCGATAGTGAGATATTATTGTA 589

Qy        137 TTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAA--A 194

Db        590 TTTTTTTTTTTTTAAAAAAGTTGAGTTAATAATAAGGTTAAAAAAAAAAAATATTAAGAGTT 649

Qy        195 ATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAAT 254

Db        650 TTTTAAATTATTATGAGGGTTTTTGTTGTTGATTATTTTATATTTTTTTTTTTTTTAAAT 709

Qy        255 TGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGG 314

Db        710 TTATAATTTGTTTTTTTTTTTTTTTTTTATGTGACGATTTTGTTTTTTGTTTTATTGAA 769

Qy        315 GGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGG 374

Db        770 GTATAAAATAGAAAATATAAAAAGAGAATTTTTATTTATTATTATTATTAAATTAATTTA 829

Qy        375 TAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGAC 434

Db        830 TTTATTTGGATGGATGGTTTATATTTTTATTTAAGTCGAGTTTTTTTTTTTACGGAGTAT 889

Qy        435 CATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTT 494

Db        890 TTATTTATTTTAAGGTTTTGTTTCGGTAATTGTTTTTGTTTTTTATATTATTAGGTTTTT 949

Qy        495 GCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAA 554

Db        950 TTTTATTAGGTTATTTTTATTAATATATAATATATAAATTTGTAGTTTTTTTATATTAAA 1009

Qy        555 AATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCT 614

Db       1010 ATTATTTTTTTTGATTTTAGGTTTATTTTTAGATATTTTTTTATTTTTGTTTTTTTTTAT 1069

Qy        615 A 615
```

Db 1070 A 1070

Search completed: May 10, 2006, 10:59:29
Job time : 557.175 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 11:00:02 ; Search time 3795.32 Seconds
(without alignments)
14608.170 Million cell updates/sec

Title: US-09-423-093B-2_COPY_858_2042
Perfect score: 1185
Sequence: 1 gtgaagtcagcggctaagtt.....gttataactgtgacttataa 1185

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

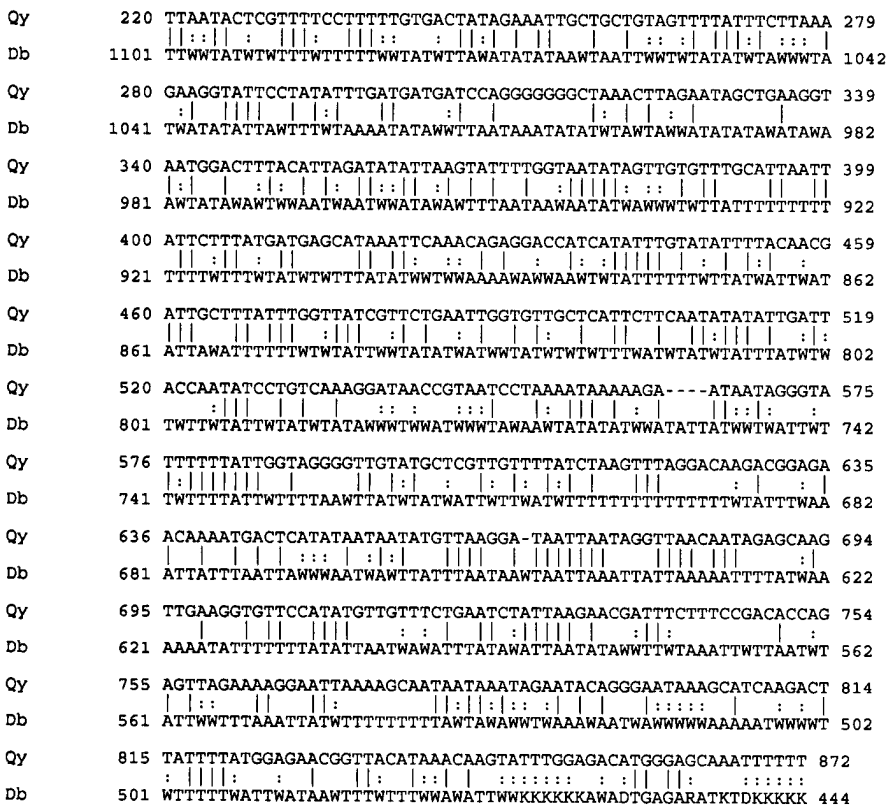
SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	99.6	8.4	1101	10	CNS00EVL	AL069706 Drosophil
2	87.8	7.4	1896	10	CG753083	CG753083 P048-1-C0
c 3	84	7.1	1101	10	CNS0039G	AL063921 Drosophil
4	80.8	6.8	1101	10	CNS00EVL	AL069706 Drosophil
5	80.4	6.8	1242	10	CL068807	CL068807 CH216-115
c 6	77.6	6.5	1101	10	CNS0145U	AL103740 Drosophil
c 7	76	6.4	1092	10	CNS020K7	AL175696 Tetraodon
8	75.8	6.4	1350	10	CG744271	CG744271 P036-4-C0
c 9	75.2	6.3	945	3	BM358135	BM358135 GA__Ea000
c 10	74.4	6.3	499	10	CW874602	CW874602 shSc2d26-
11	74.4	6.3	891	10	CW952279	CW952279 TcB38.1_B
12	74	6.2	1101	10	CNS0039G	AL063921 Drosophil
13	73.8	6.2	1101	10	CNS003DX	AL064587 Drosophil
14	73.8	6.2	1101	10	CNS0100X	AL098379 Drosophil
c 15	73	6.2	1563	9	CC206942	CC206942 CH261-191
16	72	6.1	816	10	CNS02KJV	AL201604 Tetraodon
c 17	72	6.1	1275	10	CL033318	CL033318 CH216-36F
c 18	71.8	6.1	889	10	AG362888	AG362888 Mus muscu
c 19	71.6	6.0	1116	1	AJ925982	AJ925982 AJ925982
c 20	71.4	6.0	945	11	CNS04D0K	AL285149 Tetraodon
21	71	6.0	1029	10	CNS01ZGM	AL174271 Tetraodon
22	70.6	6.0	1147	6	CD387645	CD387645 AGENCOURT
c 23	70.6	6.0	1335	10	CL646802	CL646802 CH213-123
c 24	70.6	6.0	1605	8	DN712410	DN712410 CNB02-B08
c 25	70.4	5.9	928	10	CNS00DKY	AL071865 Drosophil
26	70.4	5.9	1539	10	AG340947	AG340947 Mus muscu
c 27	70.2	5.9	994	11	CNS04NOJ	AL298972 Tetraodon
c 28	70	5.9	1199	3	BM544915	BM544915 AGENCOURT
c 29	69.8	5.9	1101	10	CNS00EO7	AL069440 Drosophil
c 30	69.6	5.9	889	10	CW973509	CW973509 AIAA-aaa5
31	69.2	5.8	928	10	CNS00DKY	AL071865 Drosophil
c 32	69	5.8	1101	10	CNS00FVE	AL071298 Drosophil
c 33	68	5.7	951	10	CW952366	CW952366 TcB38.1_E
34	68	5.7	961	7	CK230311	CK230311 ILLUMIGEN
c 35	68	5.7	1101	10	CNS003BD	AL064091 Drosophil
c 36	67.8	5.7	735	10	CW960936	CW960936 AIAA-aac1
c 37	67.8	5.7	1201	10	CNS0167M	AL106396 Drosophil
c 38	67.6	5.7	996	10	CNS00FUH	AL071063 Drosophil
c 39	67.4	5.7	826	10	AG579614	AG579614 Mus muscu
40	67.4	5.7	1027	10	CNS02T50	AL212733 Tetraodon
41	67.4	5.7	1348	10	CG749499	CG749499 P043-4-A0
c 42	67.4	5.7	1359	8	DN685273	DN685273 CGX40-G12
c 43	67.4	5.7	1451	10	CW949335	CW949335 TcB33.4_G
c 44	67.2	5.7	1277	9	CC253231	CC253231 CH261-180
c 45	67	5.7	810	10	AG486400	AG486400 Mus muscu

ALIGNMENTS

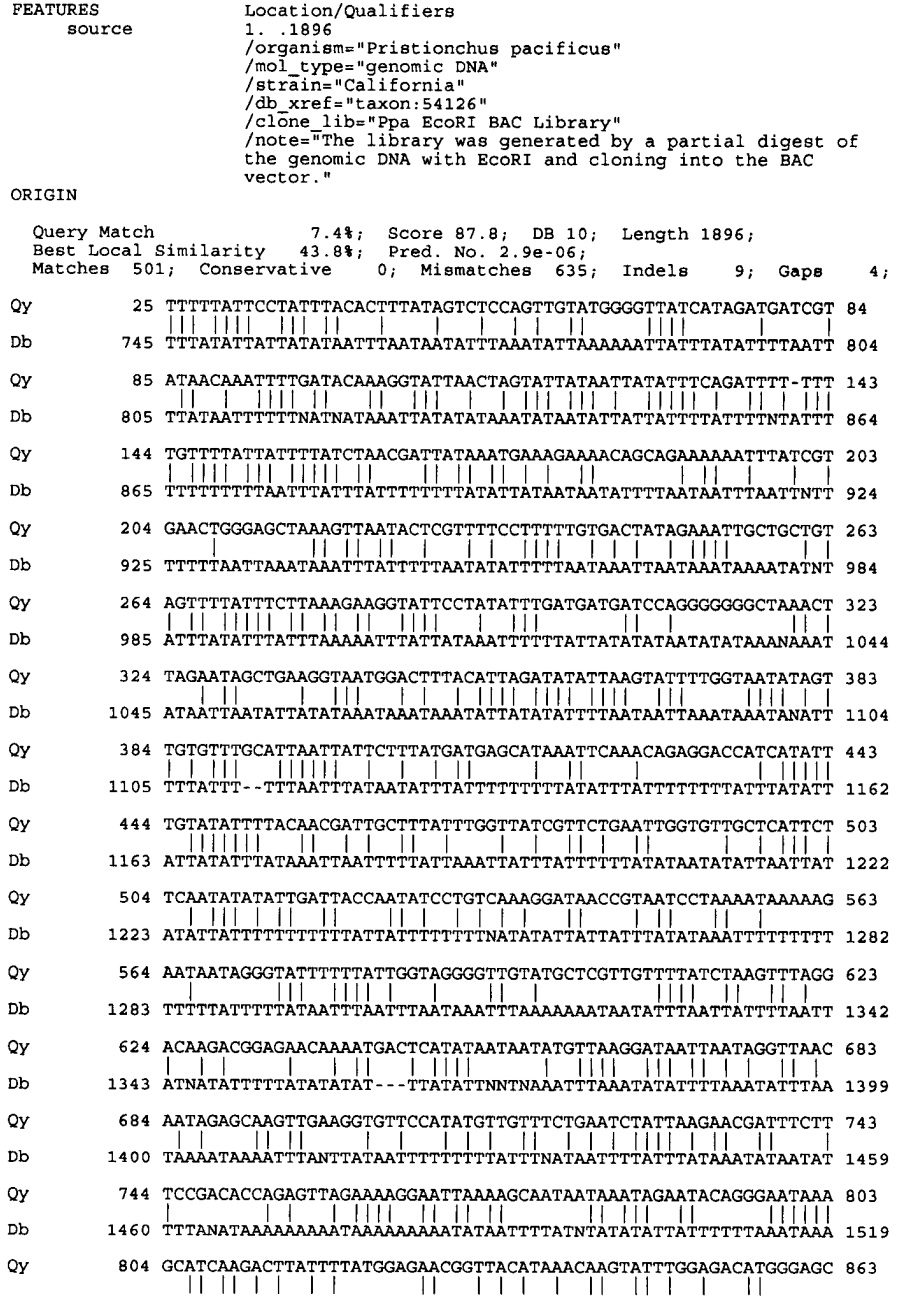
RESULT 1
CNS00EVL/c
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"
ORIGIN

Query Match 8.4%; Score 99.6; DB 10; Length 1101;
Best Local Similarity 35.6%; Pred. No. 1.9e-08;
Matches 234; Conservative 119; Mismatches 300; Indels 5; Gaps 2;



RESULT 2
CG753083
LOCUS CG753083 1896 bp DNA linear GSS 24-OCT-2003
DEFINITION P048-1-C01.za Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG753083
VERSION CG753083.1 GI:37977199
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1896)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1896
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."
ORIGIN

Query Match 7.4%; Score 87.8; DB 10; Length 1896;
Best Local Similarity 43.8%; Pred. No. 2.9e-06;
Matches 501; Conservative 0; Mismatches 635; Indels 9; Gaps 4;



```
Db       1520 ATATTAAAAATATTATATATATTAAAAATAAAAATATAAAAAATATATAATAATTATTTA 1579

Qy        864 AAATTTTTTATCAGTTACTACGTATGGAGCAGAACTGTTAGTTTTTTTTGGTTTTCTCTG 923

Db       1580 TAATNATATTTTAAATAATTATTATTATATATATAAATTNAAAATAATTATTTTTTTTNN 1639

Qy        924 TGTATTCATTATCCCTTTAGGGATATATATACCTTTTTATCTTTTAAAGAGAATGAAAAA 983

Db       1640 TAATTATATTTTATTTATTTTTAATTATAT---TNTATATATAATTTANATTTTTATATT 1696

Qy        984 AACCCATAGCTCGATAAATTGCGCATTCTATTCATATATCATTATGATTTTATTGCAATA 1043

Db       1697 TATATNAATATTTAATATATTAATAATATTTATATATTTATTTAAATTNTTATTATTTAT 1756

Qy       1044 CTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGTCCTTTTCTCTCCGTATTGATAAT 1103

Db       1757 ATATTTAATTTTTTATTTTTTAATTTTTTTATTATATATATTTATTTTTTATTANATT 1816

Qy       1104 GTGTACTCCTCTGATCTTATTGCATGATACGTTAAAGAGATTATCACGAAATGAAAATAT 1163

Db       1817 TTTTTTTATATTANATTTAATTTATTATTTTTNTAATAAAANATATTATATTTATAATAT 1876

Qy       1164 CAGTT 1168

Db       1877 TTTTT 1881
```

RESULT 3
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

Query Match 7.1%; Score 84; DB 10; Length 1101;
Best Local Similarity 17.5%; Pred. No. 1.5e-05;
Matches 140; Conservative 328; Mismatches 331; Indels 1; Gaps 1;

```
Qy        338 GTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAA 397

Db       1101 GKARRWGDDTWDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWW 1042

Qy        398 TTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCATATTTGTATATTTTACAA 457

Db       1041 WWWWWWATWDTWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADG 982

Qy        458 CGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCTTCAATATATATTGA 517

Db        981 AGRRDGGRKRDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRK 922

Qy        518 TTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAATAATAGGGTATT 577

Db        921 ADDDDGAGDKDDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWW 862

Qy        578 TTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAAC 637

Db        861 GWADADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRR 802

Qy        638 AAAATGACTCATATAATAATATG-TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTT 696

Db        801 DRKRADDKRDAADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRD 742

Qy        697 GAAGGTGTTCCATATGTTGTTTCTGAATCTATTAAGAACGATTTCTTTCCGACACCAGAG 756

Db        741 RAGTAGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRR 682

Qy        757 TTAGAAAAGGAATTAAAAGCAATAATAAATAGAATACAGGGAATAAAGCATCAAGACTTA 816

Db        681 DRATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDW 622

Qy        817 TTTTATGGAGAACGGTTACATAAACAAGTATTTGGAGACATGGGAGCAAATTTTTTATCA 876

Db        621 AKWDWKTRADRWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTT 562

Qy        877 GTTACTACGTATGGAGCAGAACTGTTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATC 936

Db        561 ATWTTWAARAAWWAWWAWATTTATWTTTWTTTWTTTTTTTTTTTTWTTTAAWWWAAWWTAT 502

Qy        937 CCTTTAGGGATATATATACCTTTTTATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCG 996

Db        501 WAAWTAAAWAAAAAAAWAATTTTTTTTTTTTWAWTAAAWWTAWTWTTTTWTTTTWWAATT 442

Qy        997 ATAAATTGCGCATTCTATTCATATATCATTATGATTTTATTGCAATACTTAGTGGCTGGG 1056

Db        441 TTTTWTWTTWWAATTATTTTWTTTWTAAAAAAAAWWWWWTWTATATKCCCCCCYCCCCCC 382

Qy       1057 AATGCATCGGCCTTCTTTTTTGGTCCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTG 1116

Db        381 YCCYCCCCTCCYYCCTCCYTCYCCCYYYYYYYYYYYYTYCCCCTYYYYYYCCCTTTYYYY 322

Qy       1117 ATCTTATTGCATGATACGTT 1136

Db        321 YYYCYYYYCCYTCNCNCCYY 302
```

RESULT 4
CNS00EVL
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

```
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             6.8%;  Score 80.8;  DB 10;  Length 1101;
  Best Local Similarity   33.6%;  Pred. No. 5.9e-05;
  Matches  206;  Conservative  117;  Mismatches  289;  Indels    2;  Gaps    2;

Qy         87 AACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTTTTTTTGT 146
              :::::::  |  :: ||:   :|||   | :|  | :|::::||||   ::::|: || |
Db        465 MMMMMWWAATWTWWAAAWAAAWTTATWAATWAAAAAWAWWWWATTTTTWWWWWTWATTWT 524

Qy        147 TTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTATCGTGAA 206
              ||:|:: :| :    ||  | :|||| | ||| :||  :    |:||  | :|   |  |
Db        525 TTWAWWTWTAWTAAAAAAAAAWATAATTTAAAWWAATAWATTAAWAATTTAWAAWWTATA 584

Qy        207 CTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGTAGT 266
              |   :  |||| :|:||      |       | | |   :||| || ||        |
Db        585 TTAATWTATAAATWTWATTAATATAAAAAAATATTTTTTWATAAAATTTTTAATAATTTA 644

Qy        267 TTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTAAACTTAG 326
              |||:|| ||||| ||  :|||  :  |||  || ||  :  |          |||   |
Db        645 ATTAWTTATTAAATAAWTWATTWWWTAATTAAATAATTTWAAATAWAAAAAAAAAAAAAA 704

Qy        327 A-ATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATATAGTTG 385
              | ||:  : |   ||:   | || | |: | | |  :  :| |: ::     ||::|  |
Db        705 AWATWAAWAATWATAWATAAWTTAAAAWAATAAAAWAAWAATWAWWATAATATWWATATA 764

Qy        386 TGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCATATTTG 445
              | :||  |:::||:: : :|    || :    :||:  |:| | |  :  |:  :|  :
Db        765 TAWTTWTAWWWATWWAWWWTATAWATAWAATAWAAWAWAWATAAATAWATAWATWAAAWA 824

Qy        446 TATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCTTC 505
              :|:|| ::| :|    :   | : :     ||       :| |  |  :    | | :
Db        825 WAWATAWWATWATATAWWAATAWAWAAAAAATWTAATATWAATWATAAWAAAAAATAWAW 884

Qy        506 AATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAA 565
                ::|:| || : ::  |  |   : | :||  | ||    ||   |||:|::: :| |
Db        885 TTWWTWTTTTWWAWWATATAAAWAWATAWAAAWAAAAAAAAAAAAAATAAWAWWWTWATAT 944

Qy        566 TAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGTTTAGGAC 625
              |: ||    |:|:| | ::  |:    :  |:|     |:  |:||| ||  :| :  :
Db        945 TWTTATTAAAWTWTAT-WWATTWATTWWAWTWTATAWTTWTATWTATATATWWTAWTAWA 1003

Qy        626 AAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAATTAATAGGTTAACAA 685
```

```
                | |   |  | :: ||   | |:| |:|||||| |:|   :: |:  |||  ::||  |
Db       1004 TATATTTATTAAWWTATATTTTAWAAAWTAATATATWATAWWWTAWATATAWAWWAATTA 1063

Qy        686 TAGAGCAAGTTGAA 699
              :  |   | :| |:
Db       1064 WTTATATATWTAAW 1077


RESULT 5
CL068807
LOCUS       CL068807               1242 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-115B3_Sp6.1 CH216 Xenopus tropicalis genomic clone
            CH216-115B3, genomic survey sequence.
ACCESSION   CL068807
VERSION     CL068807.1  GI:40524720
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1242)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 7
            High quality sequence stop: 57.
FEATURES             Location/Qualifiers
     source          1. .1242
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-115B3"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             6.8%;  Score 80.4;  DB 10;  Length 1242;
  Best Local Similarity   46.3%;  Pred. No. 7e-05;
  Matches  467;  Conservative    0;  Mismatches  511;  Indels   30;  Gaps    5;

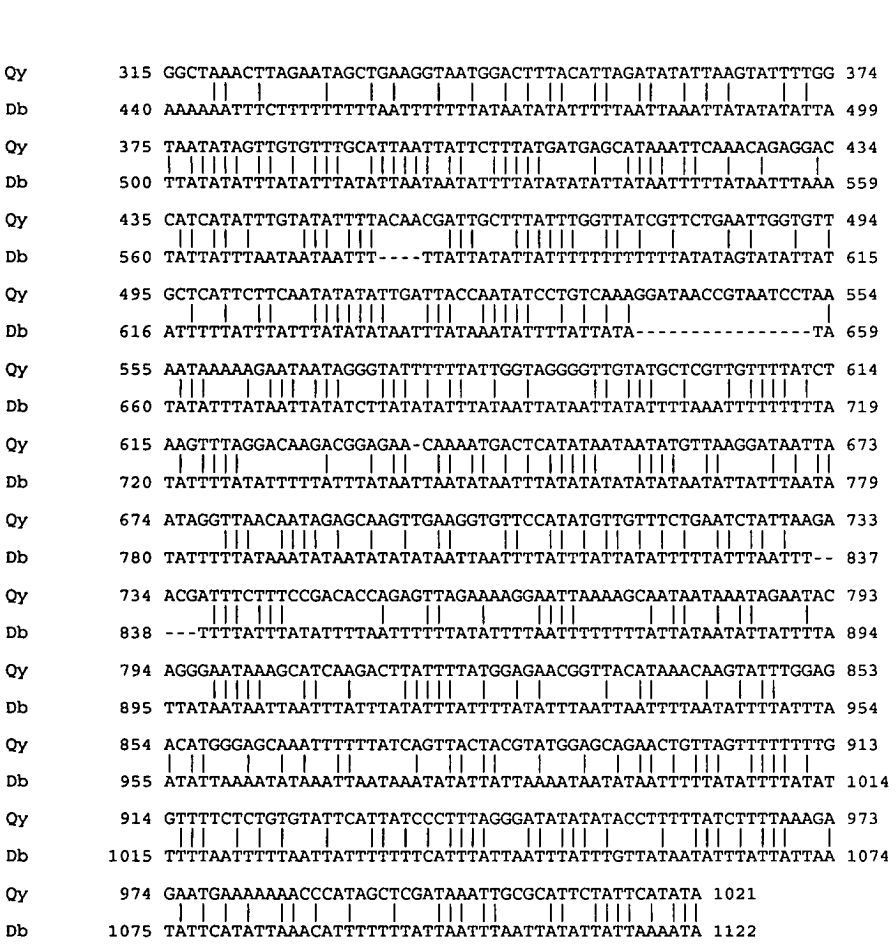
Qy         15 TAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCAT 74
              | | ||   | ||||||   |||||| | ||| |  | |  | || |||    |||  ||
Db        144 TTATTTTTATATTTTATATATATTTAAAATTTTTTATTTATATTTTTATTATATTAATAT 203

Qy         75 AGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCA 134
              ||  |  ||     |  || |      |       ||| |   ||| |||||   ||| |
Db        204 TTATATTATTAATTTATTTTATATATTATTTTATATAATTTTAATTTTAATTTATTTTTA 263

Qy        135 GATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAA 194
              ||||||||||||  | |   |||||       ||||||  |  |  | |        |  |
Db        264 TATTTTTTTTGTTATTTATATTTATAATTTAATTATATTTATATTAGATATTTTTATTTA 323

Qy        195 ATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAAT 254
              ||||||  |  | |    |  | |    |  || |  ||||  ||||  | |  ||  | |
Db        324 ATTTATTATATATTTATAATTTATTATTTTAATATTTTTTATTTTAATTATATTAATATA 383

Qy        255 TGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGG 314
              |    | |  |||    | |||||  |  ||||  ||||  ||  |  |    |
Db        384 T----ATTTTTTTTATAATATTAAAATAATTATTTTTATAATTATTTTTTTAAAAAAAAA 439
```



RESULT 6
CNS0145U/c
LOCUS CNS0145U 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN11O16 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL103740
VERSION AL103740.1 GI:5615351
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN11O16"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"
ORIGIN

Query Match 6.5%; Score 77.6; DB 10; Length 1101;
Best Local Similarity 33.3%; Pred. No. 0.00023;
Matches 256; Conservative 77; Mismatches 433; Indels 2; Gaps 1;

RESULT 7

```
CNS020K7/c
LOCUS       CNS020K7                1092 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            222L11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL175696
VERSION     AL175696.1  GI:7813753
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1092)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="222L11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG222CF06LP1
                     end : T7"
ORIGIN

  Query Match             6.4%;  Score 76;  DB 10;  Length 1092;
  Best Local Similarity   39.4%;  Pred. No. 0.00047;
  Matches  218;  Conservative   70;  Mismatches  253;  Indels   13;  Gaps    2;

Qy         24 TTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATAGATGATCG 83
Db       1092 TTTAATTTTTTTTTTTTTTTTTTTTTTTATTTTAWTTTTTAATTTATTTTTTTTTTTTAA 1033

Qy         84 TATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTTTTTT 143
Db       1032 TATATAAATATTTTATTTWTTTATATTTTATTTTTTTTTTTTTTTAAAAATAAATTAAWW 973

Qy        144 TGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTATCGT 203
Db        972 TTTTNWTTTATATTTTTTTTTTTTTWTTAWAAAAAWWAAAAAAAAAAAAAAWATTTAWAWW 913

Qy        204 GAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGT 263
Db        912 TWA--------WAAAAAAAAWWAWTTWTTTTTTAATTTTTTTTTTTTAWATTWWAATTTWT 861

Qy        264 AGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTAAACT 323
```

```
Db        860 AAAWTWTTTTAAAWAATTATTTWWTTTTTTAWTAAWTAAAWTNTTTTWTTTAAAATAWAA 801

Qy        324 TAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATATAGT 383
Db        800 WWTWATWWWTAAAAWTAAWAWWWTTTWTTTTWWWWWWWTWAAAAAAWAAWTWWWTTWTTW 741

Qy        384 TGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCATATT 443
Db        740 TTTTTTWWTTWATTWAWTTTTTTTTTT-----TTTTTTTAAAAAAAAAAAAAAAAAAAA 686

Qy        444 TGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCT 503
Db        685 TTTTTTTTTTAAWWWWWWWWTTWTTTTTTTWTAAAAAAAAWWTTTTTTTTTTTTTWWWWT 626

Qy        504 TCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAG 563
Db        625 TAAAWATAAAWTTTWAAAAWDTTTKKGGSSSSSGGGTTGGCGGAGTTTTTGYTGAARAAG 566

Qy        564 AATAATAGGGTATT 577
Db        565 AACGACWACGGCTT 552


RESULT 8
CG744271
LOCUS       CG744271                1350 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P036-4-C03.ya Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG744271
VERSION     CG744271.1  GI:37965139
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1350)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1350
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN

  Query Match             6.4%;  Score 75.8;  DB 10;  Length 1350;
  Best Local Similarity   47.6%;  Pred. No. 0.0005;
  Matches  273;  Conservative    0;  Mismatches  296;  Indels    5;  Gaps    2;

Qy         15 TAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCAT 74
Db        596 TATTTTATTTTTATTTTTTTTTTTTATTTATATTTATATTTATATTATTTTTAATATATT 655

Qy         75 AGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCA 134
Db        656 TATTTATAATTATTTTTATTTTTTATATTTTTTATTAAATTTAATTTTAATTATATTTAA 715
```

```
Qy        135 -GATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAA 193
              |||||||| |||| || |||||| ||     ||||  | | | ||           |
Db        716 TAATTTTTTTTATTTTTTTTTTTATTTATAATATATATTTTATATAATTATTTTATTTA 775

Qy        194 AATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAA 253
              ||  | |     | |    | || || ||| |   |||  ||||| |   |||  | |
Db        776 AANATTTTTAANATTAATTAATTAATTTTATAATAAATTTATTTTTTTTTTTTTATTTATA 835

Qy        254 TTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGG 313
              |   |     | | |||| |   |    |   ||     ||||||| || || ||  |
Db        836 TANATATATTTTTATAATTTTTATATATAAATAAATATATATTTTATAATAATTTAAATA 895

Qy        314 GGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTG 373
              |      || ||    | ||  ||||    |||| | || ||||||||  |  | |
Db        896 TTAATTTTAATAAAAATTATTAAATTAATTTTTTTTAAAATATATATATTATATTATATT 955

Qy        374 GTAATATAGTT----GTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAG 429
              || |||| ||     |  ||   |||||||| | | | || |  ||| | |
Db        956 TTATTATATTTTTTATTAATTAATTTAATTATTTTATTTTATAATTATATAATTTTNTTT 1015

Qy        430 AGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTG 489
              |  |   |  |  |  |  ||||||     ||   | |   |  | |   |    |||
Db       1016 ATTATTTTTTTTATAATTTATTTTTATTTAATAATTATTAATATTATTATATTATTATTT 1075

Qy        490 GTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAAT 549
              | ||  |  || ||   | |    | | | | | |  |    ||     |||
Db       1076 ATTTTTTTNNTTATTTTTTTTTNTATTTTNATTATTTTTTTAATATTTTTTATTTTAAAT 1135

Qy        550 CCTAAAATAAAAAGAATAATAGGGTATTTTTTAT 583
                |    || | |   ||||    | | ||||||
Db       1136 ATTTTTTTATATANTTTAATTTTTTTTATTTTAT 1169
```

RESULT 9
BM358135/c
LOCUS BM358135 945 bp mRNA linear EST 09-JAN-2002
DEFINITION GA__Ea0005I15r Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum cDNA clone GA__Ea0005I15r, mRNA sequence.
ACCESSION BM358135
VERSION BM358135.1 GI:18098881
KEYWORDS EST.
SOURCE Gossypium arboreum
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE 1 (bases 1 to 945)
AUTHORS Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber
JOURNAL Unpublished (2000)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total High Quality bases = 126
Seq primer: TAATACGACTCACTATAGGG
High quality sequence start: 3
High quality sequence stop: 842.
FEATURES Location/Qualifiers
source 1. .945
/organism="Gossypium arboreum"
/mol_type="mRNA"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA__Ea0005I15r"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Query Match 6.3%; Score 75.2; DB 3; Length 945;
Best Local Similarity 46.7%; Pred. No. 0.00066;
Matches 256; Conservative 0; Mismatches 291; Indels 1; Gaps 1;

```
Qy         42 ACTTTATAGTCTCCAGTTGTATGGGGTTATCATAGATGATCGTATAACAAATTTTGATAC 101
              | |  |||       | || ||     ||||     || ||   ||   |  | ||
Db        877 ATTAAATAAAANAAATTTATAATATTTTATTTATAATAATATAATTTTATTTATTATATT 818

Qy        102 AAAGGTATTAACTAGTATTATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCT 161
              |||  || | |  |     | ||  ||||| | |  | || |||   ||||| ||| || |
Db        817 AAATATAATTATAATATATTTATATATATTATTTATTATTTTAATTATTTTTAATAT 758

Qy        162 AACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTT 221
              |   | ||| |||   | | || |      |||  ||||  | || |   |  |||  ||
Db        757 ATTAAATATTAATTTTANANAAAAAATATTAAATATTATTATTAAATTTTATTTAATTTT 698

Qy        222 AATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGA 281
              |      | ||  || || |  | |   | ||      | || || | ||| || | ||
Db        697 TTTTTATATATTTTTTATTTTATTTTTTTTATTTTTATTTATATTTATTTTTTTTTATGA 638

Qy        282 AGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAA 341
              ||    |  | |  |  | | |      | |||   ||   |   |    |  |
Db        637 TTTATTTTTATTTATATAANTTATTTGATAAAAAGATAATTATAATTTTATTTTTTTTTT 578

Qy        342 TGGACTTTACATTAGATATATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTAT 401
              |    |||   ||  || | ||  || ||||  |   |    || | |||   ||  ||||
Db        577 TTTTTTTTTTTTTTTATTTTTTTGTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTAT 518

Qy        402 TCTTTATGATGAGCATAAATTCAAACAGAGGACCATCATAT-TTGTATATTTTACAACGA 460
              | |  |  |      |   ||   |          |  | | || | | || |      |
Db        517 TTATATTATTTTTTTTTTTTTTTTATTTTTTTATTTATTTTGTTTTTTTTTATTATTTTA 458

Qy        461 TTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTA 520
              || ||||||   || | ||| | |||   | ||  | |||  |  || |||||   ||
Db        457 ATGTTTTATTAATTTTTTGTTGTTAATGTTTTTTTTTTTATTTATTTATTTATATAATTTT 398

Qy        521 CCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTT 580
              || | |   |  |||  | ||   |||  ||||||||  | ||  |     |||||||
Db        397 TCATTTTATTNNNAAAANAAAAAAAAAATAATAAAATAATTAAAAATAAGTTTTATTTTT 338

Qy        581 TATTGGTA 588
                ||  ||
Db        337 AATAATTA 330
```

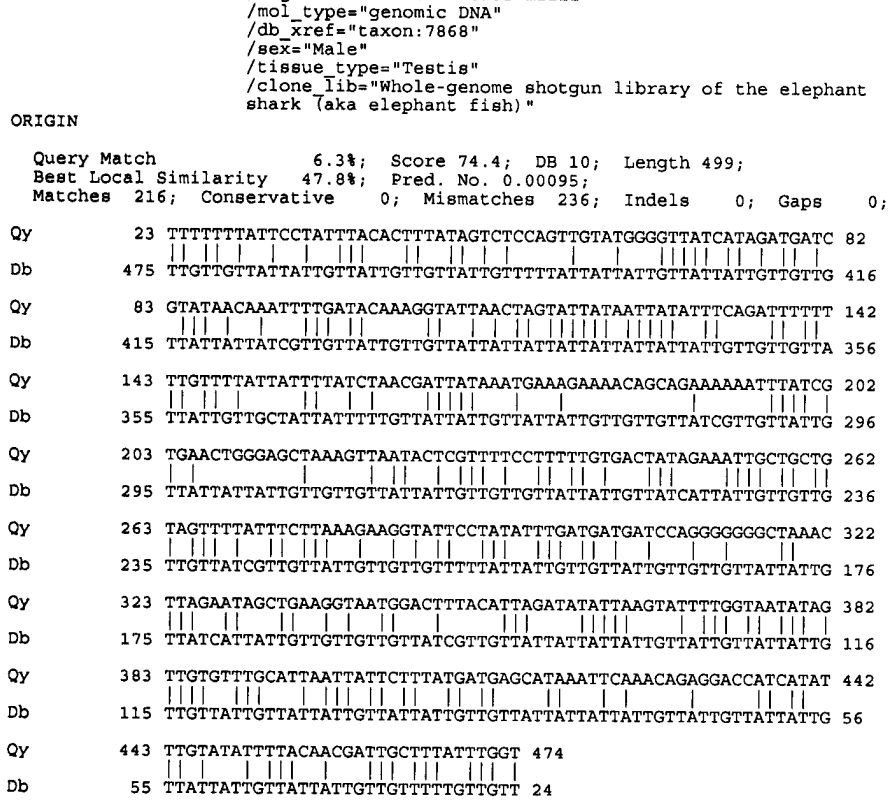
RESULT 10
CW874602/c
LOCUS CW874602 499 bp DNA linear GSS 12-FEB-2005
DEFINITION shSc2d26-44.g_047.ab1 Whole-genome shotgun library of the elephant shark (aka elephant fish) Callorhinchus milii genomic, genomic survey sequence.
ACCESSION CW874602
VERSION CW874602.1 GI:59701237
KEYWORDS GSS.
SOURCE Callorhinchus milii (elephantfish)
ORGANISM Callorhinchus milii
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes; Holocephali; Chimaeriformes; Callorhinchidae; Callorhinchus.
REFERENCE 1 (bases 1 to 499)
AUTHORS Venkatesh,B., Tay,A., Dandona,N., Patil,J.G. and Brenner,S.
TITLE A compact cartilaginous fish model genome
JOURNAL Curr. Biol. 15 (3), R82-R83 (2005)
PUBMED 15694293
COMMENT Contact: Venkatesh B
Molecular Genetics Lab
Institute of Molecular and Cell Biology
61 Biopolis Drive, Singapore 138673

Tel: 65 6586 9571
Fax: 65 6779 1117
Email: mcbbv@imcb.a-star.edu.sg
Whole-genome shotgun sequences of the elephant shark (aka elephant fish)
Class: shotgun.

FEATURES Location/Qualifiers
source 1. .499
/organism="Callorhinchus milii"
/mol_type="genomic DNA"
/db_xref="taxon:7868"
/sex="Male"
/tissue_type="Testis"
/clone_lib="Whole-genome shotgun library of the elephant shark (aka elephant fish)"

ORIGIN

Query Match 6.3%; Score 74.4; DB 10; Length 499;
Best Local Similarity 47.8%; Pred. No. 0.00095;
Matches 216; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

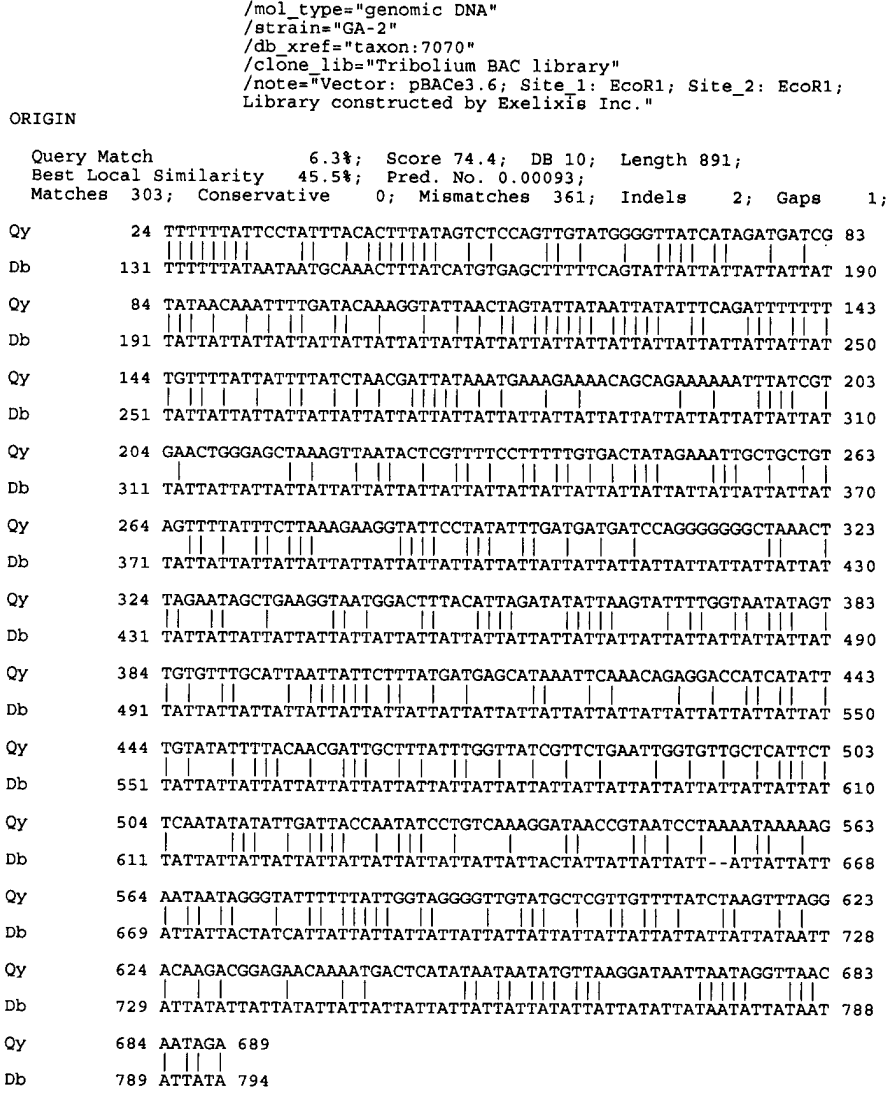


RESULT 11
CW952279
LOCUS CW952279 891 bp DNA linear GSS 20-DEC-2004
DEFINITION TcB38.1_B02_T7 Tribolium BAC library Tribolium castaneum genomic, genomic survey sequence.
ACCESSION CW952279
VERSION CW952279.1 GI:56735316
KEYWORDS GSS.
SOURCE Tribolium castaneum (red flour beetle)
ORGANISM Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tribolium.
REFERENCE 1 (bases 1 to 891)
AUTHORS Savard,J. and Tautz,D.
TITLE Tribolium castaneum BAC-ends sequencing project
JOURNAL Unpublished (2003)
COMMENT Contact: Savard, J.
Abteilung fur Evolutionsgenetik, AG Tautz
Institut fur Genetik, Universitat zu Koln
Weyertal 121, 50931 Koln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: savard@uni-koeln.de
Class: BAC ends.

FEATURES Location/Qualifiers
source 1. .891
/organism="Tribolium castaneum"
/mol_type="genomic DNA"
/strain="GA-2"
/db_xref="taxon:7070"
/clone_lib="Tribolium BAC library"
/note="Vector: pBACe3.6; Site_1: EcoR1; Site_2: EcoR1; Library constructed by Exelixis Inc."

ORIGIN

Query Match 6.3%; Score 74.4; DB 10; Length 891;
Best Local Similarity 45.5%; Pred. No. 0.00093;
Matches 303; Conservative 0; Mismatches 361; Indels 2; Gaps 1;



RESULT 12
CNS0039G
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999

```
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             6.2%;  Score 74;  DB 10;  Length 1101;
  Best Local Similarity   19.3%;  Pred. No. 0.0011;
  Matches  127;  Conservative  263;  Mismatches  264;  Indels    4;  Gaps    1;

Qy          77 ATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGA 136
               ||  ::  :|:|| |||||:: | | |||   :| :: |  : : | ||  | |  | :
Db         428 ATTWWAAWAWAAAAAATTWWAAAAWAAAAWAWTAWWTTTAWTWAAAAAAAAAAAATTWTT 487

Qy         137 TTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAAT 196
               |||||:||  :||: ||::||:: | |  :  | |||  ||: ||:| |    || : ::
Db         488 TTTTTWTTTAWTTWATAWWTTWWWTTAAAWAAAAAAAAAAAAWAAAWAAAWATAAATWTWW 547

Qy         197 TTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTG 256
               |::|  | :| :   | : :|| ::: :  :  :::|: |:| | : :   |: :|:||
Db         548 TWWTTYTTWAAWATAAAMCMAAWYYHTYTTYTYHYYTYWTYTMTWHYHTMYTHAWAHTTW 607

Qy         257 CTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGG 316
               :: :: :| :::  :::  :|:: :   || :::: |  :: :: :: : : :
Db         608 YHWYHTYAMWHWMTWHWTMWAWWHWTTYTAAYYYYYTCMYYYHYMHWHHAHAHAAAWWTT 667

Qy         317 CTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTA 376
               :|:: : :: | :: :: :    :  : :: : :::  ||::|: : :  :::  :
Db         668 HTWWTHAYHWATYHYYYYMYCAMMCMCTHTCHHCYYYYHHYTAHHTHTHHWYAHYYMWYY 727

Qy         377 ATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCA 436
               ::|:   : |  |: ::::::::: ::||:   |: : ::::|::  ||| |    |
Db         728 MWAYYWMYCTACTYHYHHHHHYHWAYHTTWYAWAHAMWMWHHAHYAAAAAWA----AWAT 783

Qy         437 TCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGC 496
               |:::::|| ::::|: ::: ::   : ||:: :::  |  || |: : :::|  : ::
Db         784 THHYHHTTHYMHHTYMYHYYMYTCCYMCTYHCWHYYHTAYTCWTWTHHWMWTWTHWYHHT 843

Qy         497 TCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAA 556
               : ::| |: |:::|:|: : :::|  : :||:|:  : ::  :::::: ::: |: :::
Db         844 WWHHTTTHWAWWHTHTWCWWWWHATTWTWATHCWACMTMHWHHWMHMHHHHMACHAHHHT 903

Qy         557 TAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTAA 616
               :: : : :::   :   |::: | :: :  ::    ::   :  |: |      : :::
Db         904 HMCMCHHHMHCTCHHHHTMYHMTCHWMWHMHWHWHHWWMATWMTTMTTMMMMCCMMHHHC 963

Qy         617 GTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAATTAATA 676
               :::  :::   :    : :|  : :: |   :  |::| :  :::    :| | ||:
Db         964 HMYHMMHMYMYCCHYYCTCHTHATTHYHYMCTCYHYCTWHTYWTAYWWAWTAHAMTTATW 1023

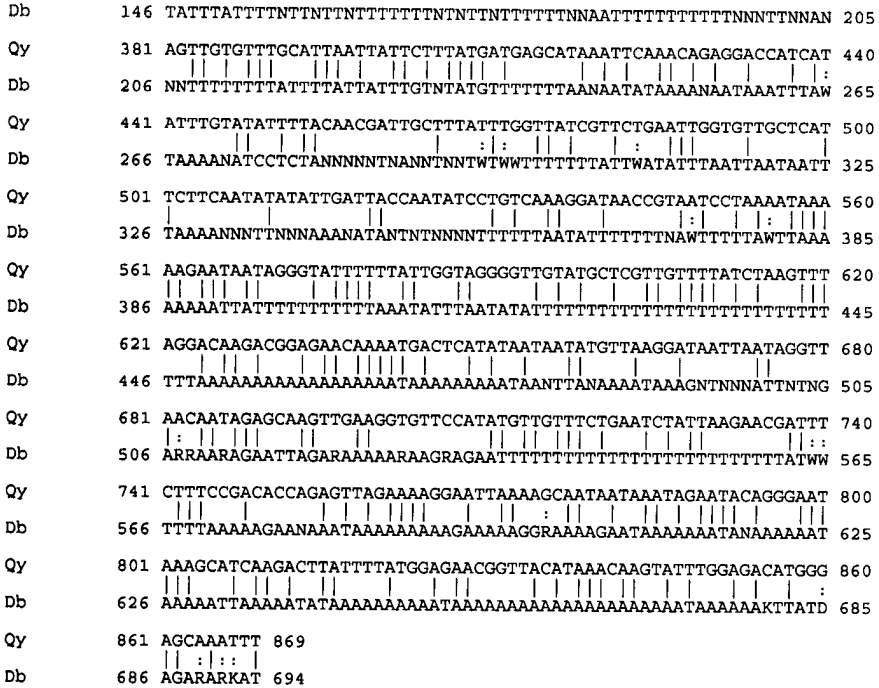
Qy         677 GGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCTATTAAGAA 734
               : :: |::| :  :: ::  |    | :::|:: |: |::::   :::: ::: ::
Db        1024 WWWMHWWAHWATWWWWWWWWATAWWACTCHHWTWYHTHCTWYYHHTYHMWWAWWMAWH 1081


RESULT 13
CNS003DX
LOCUS       CNS003DX                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR08I09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL064587
VERSION     AL064587.1  GI:4941939
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08I09"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             6.2%;  Score 73.8;  DB 10;  Length 1101;
  Best Local Similarity   39.6%;  Pred. No. 0.0012;
  Matches  241;  Conservative   14;  Mismatches  354;  Indels    0;  Gaps    0;

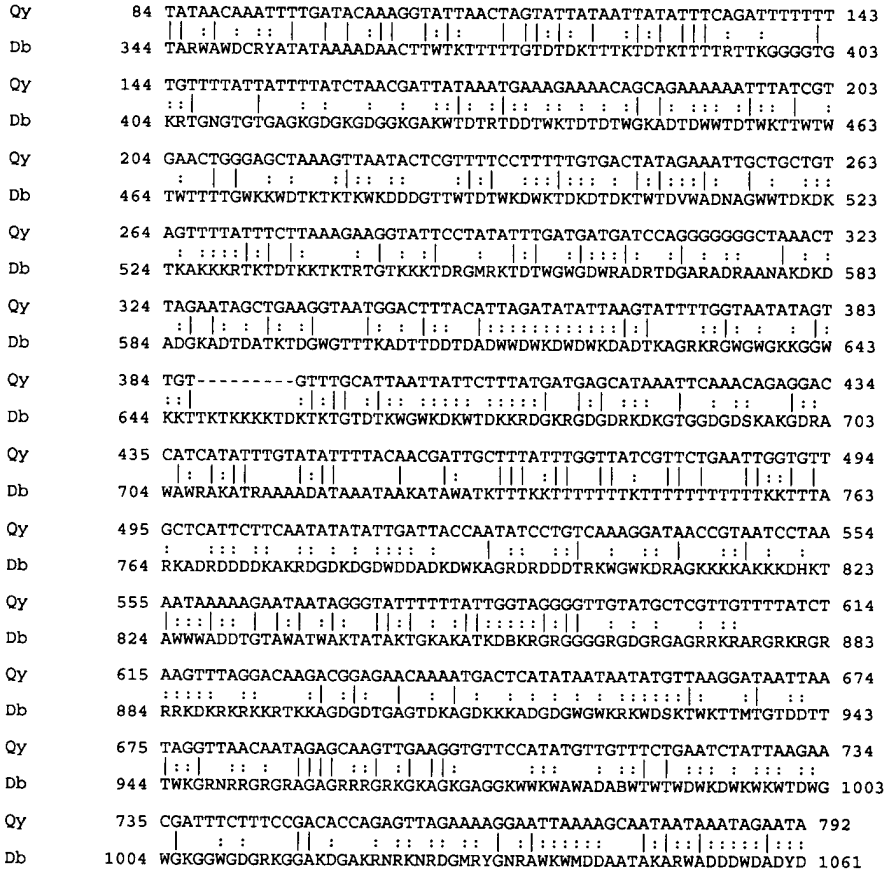
Qy         261 TGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTAA 320
               | |  || | ||  ||  |  |  | ||  | | ||| | ||   |  | |      | |
Db          86 TNTTNTTNTNTTNTTTNTATTATMTGTTATTNTNTTTTAAGAATTTATATGTATATATTA 145

Qy         321 ACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATAT 380
                  ||    |   |     |  |     ||   ||    |  ||   | |||   |   |
```



RESULT 14
CNS0100X
LOCUS CNS0100X 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL098379
VERSION AL098379.1 GI:5609990
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN03G04"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : SP6"
ORIGIN

Query Match 6.2%; Score 73.8; DB 10; Length 1101;
Best Local Similarity 19.4%; Pred. No. 0.0012;
Matches 139; Conservative 277; Mismatches 293; Indels 9; Gaps 1;



RESULT 15
CC206942/c
LOCUS CC206942 1563 bp DNA linear GSS 09-MAY-2003
DEFINITION CH261-191L18_Sp6.1 CH261 Gallus gallus genomic clone CH261-191L18, genomic survey sequence.
ACCESSION CC206942
VERSION CC206942.1 GI:30491401
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1563)
AUTHORS Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,

```
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 301
            High quality sequence stop: 608.
FEATURES             Location/Qualifiers
     source          1. .1563
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-191L18"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN

  Query Match             6.2%;  Score 73;  DB 9;  Length 1563;
  Best Local Similarity   47.3%;  Pred. No. 0.0017;
  Matches  252;  Conservative    0;  Mismatches  280;  Indels    1;  Gaps    1;

Qy        102 AAAGGTATTAACTAGTATTATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCT 161
              ||   | || | ||  ||| | ||||||  |   ||||| ||  || || | |||||| |
Db       1219 AATATTTTTTATTAAAATTTTTATTATAAATATTATTTTATTAATTATAATTTTTTATAT 1160

Qy        162 AACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTT 221
                 ||| || ||      | || |    | | |  |  | |   |   |       ||
Db       1159 TTTAATTTTATATTTGTAATAATATATAATATATATATTAATTTTTTATTAAAATTTTTT 1100

Qy        222 AATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGA 281
                |  |     |  || |    | | ||  | ||  |  | |   |||| || | ||
Db       1099 TTTTTTACAAATAATTATAAATAATTTATTATTTTTTTTTTATGAATTTAATTTTAAATTT 1040

Qy        282 AGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAA 341
              |  ||||  ||| |||  |  |      |         |||| ||           |    |
Db       1039 AAATATTTTTATTTTTTTTATTTTAATATTAATATTTAAATTTTTTTATATAATATTATA 980

Qy        342 TGGACTTTACATTAGATATATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTAT 401
              |  | |||   ||  ||| || |    || |  || || |  | | |||  || | | |
Db        979 TAAAATTTTATTTGTATAAATAATTATTTATTTTATTAAAAATTTATTTAAATAATTAAA 920

Qy        402 TCTTTA-TGATGAGCATAAATTCAAACAGAGGACCATCATATTTGTATATTTTACAACGA 460
              | |||| |  | |  | |  ||  || | |      ||  | |||  ||| |||| ||  |
Db        919 TTTTTATTTTTTATTAAATTTTTTAAGAAATTTAAATTTTTTTTTAATAATTTAAAATTA 860

Qy        461 TTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTA 520
              |     ||| |   | | ||  | || |  | ||  | ||  || ||||||  || ||||
Db        859 TAATAATATATTAGTTTAGTATTAAAATTTTTTTTATAATCATTAAATATAATTTAATTA 800

Qy        521 CCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTT 580
                |      | |   |     |||  | |   | ||  || |||    | ||| |
Db        799 ATTTGAAATAAATATATTTTTTTTTAAATATTGAGTAATAATAAAGAATTTCTTGTTTGT 740

Qy        581 TATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGTTTAGGACAAGACGGA 633
              |||| |||       ||    || |  | | ||| ||| ||| | |   |  ||
Db        739 TATTAGTATTTTAGGTTATGTCATGATGTAATCAAAGGTTATGTCTGCATTGA 687


Search completed: May 11, 2006, 05:00:43
Job time : 3798.32 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2006, 20:34:42 ; Search time 165.291 Seconds
(without alignments)
12743.667 Million cell updates/sec

Title: US-09-423-093B-2_COPY_858_2042
Perfect score: 1185
Sequence: 1 gtgaagtcagcggctaagtt.....gttataactgtgacttataa 1185

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	1183.4	99.9	14187	3	US-09-453-702B-121	Sequence 121, App
c	2	1183.4	99.9	14187	3	US-10-114-170-121	Sequence 121, App
	3	77.4	6.5	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	4	66.2	5.6	18773	3	US-09-949-016-14164	Sequence 14164, A
c	5	64.2	5.4	205044	3	US-09-949-016-15851	Sequence 15851, A
c	6	64.2	5.4	205044	3	US-09-949-016-15852	Sequence 15852, A
c	7	64.2	5.4	205044	3	US-09-949-016-15853	Sequence 15853, A
c	8	64.2	5.4	223471	3	US-09-949-016-12387	Sequence 12387, A
c	9	64.2	5.4	223471	3	US-09-949-016-12724	Sequence 12724, A
c	10	64.2	5.4	223471	3	US-09-949-016-12725	Sequence 12725, A
c	11	58.6	4.9	1141	3	US-09-806-708B-22	Sequence 22, Appl
	12	57.2	4.8	601	3	US-09-949-016-30530	Sequence 30530, A
	13	57.2	4.8	601	3	US-09-949-016-30531	Sequence 30531, A
	14	57.2	4.8	601	3	US-09-949-016-37149	Sequence 37149, A
	15	57.2	4.8	601	3	US-09-949-016-37150	Sequence 37150, A
	16	57.2	4.8	601	3	US-09-949-016-37163	Sequence 37163, A
	17	57.2	4.8	601	3	US-09-949-016-37164	Sequence 37164, A
	18	57.2	4.8	601	3	US-09-949-016-145867	Sequence 145867,
	19	57.2	4.8	601	3	US-09-949-016-145868	Sequence 145868,
	20	57.2	4.8	601	3	US-09-949-016-146135	Sequence 146135,
	21	57.2	4.8	601	3	US-09-949-016-146136	Sequence 146136,
	22	57.2	4.8	601	3	US-09-949-016-146403	Sequence 146403,
	23	57.2	4.8	601	3	US-09-949-016-146404	Sequence 146404,
c	24	55.8	4.7	187169	3	US-09-949-016-12776	Sequence 12776, A
c	25	55.8	4.7	191569	3	US-09-949-016-15940	Sequence 15940, A
c	26	54.8	4.6	32392	3	US-09-662-254B-27	Sequence 27, Appl
c	27	54.6	4.6	700	3	US-09-735-271-1038	Sequence 1038, Ap
c	28	53.8	4.5	200663	3	US-09-949-016-12569	Sequence 12569, A
c	29	53.8	4.5	203093	3	US-09-949-016-14445	Sequence 14445, A
c	30	53.2	4.5	95255	3	US-09-949-016-17067	Sequence 17067, A
c	31	53	4.5	1864	3	US-09-468-265-4	Sequence 4, Appli
	32	53	4.5	18773	3	US-09-949-016-14164	Sequence 14164, A
	33	53	4.5	205044	3	US-09-949-016-15851	Sequence 15851, A
	34	53	4.5	205044	3	US-09-949-016-15852	Sequence 15852, A
	35	53	4.5	205044	3	US-09-949-016-15853	Sequence 15853, A
	36	53	4.5	223471	3	US-09-949-016-12387	Sequence 12387, A
	37	53	4.5	223471	3	US-09-949-016-12724	Sequence 12724, A
	38	53	4.5	223471	3	US-09-949-016-12725	Sequence 12725, A
c	39	53	4.5	258775	3	US-09-949-016-16435	Sequence 16435, A
c	40	52.8	4.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
	41	52.6	4.4	832	3	US-09-621-976-2813	Sequence 2813, Ap
c	42	52.4	4.4	1132	3	US-09-533-559-5452	Sequence 5452, Ap
c	43	52.2	4.4	601	3	US-09-949-016-93907	Sequence 93907, A
	44	52.2	4.4	50000	3	US-09-662-254B-23	Sequence 23, Appl
c	45	52.2	4.4	50000	3	US-09-662-254B-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-09-453-702B-121/c
; Sequence 121, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14187
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-09-453-702B-121

```
  Query Match             99.9%;  Score 1183.4;  DB 3;  Length 14187;
  Best Local Similarity   99.9%;  Pred. No. 1.9e-252;
  Matches 1184;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13143 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 13084

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13083 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 13024

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13023 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 12964

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12963 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 12904

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12903 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 12844

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12843 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 12784

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12783 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 12724

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12723 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 12664

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12663 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 12604

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12603 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 12544

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12543 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 12484

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      12483 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAAGTGTTCCATATGTTGTTTCT 12424

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12423 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 12364

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12363 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 12304

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12303 CAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTG 12244

Qy        901 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12243 TTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 12184

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12183 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 12124
```

```
Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12123 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 12064

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12063 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 12004

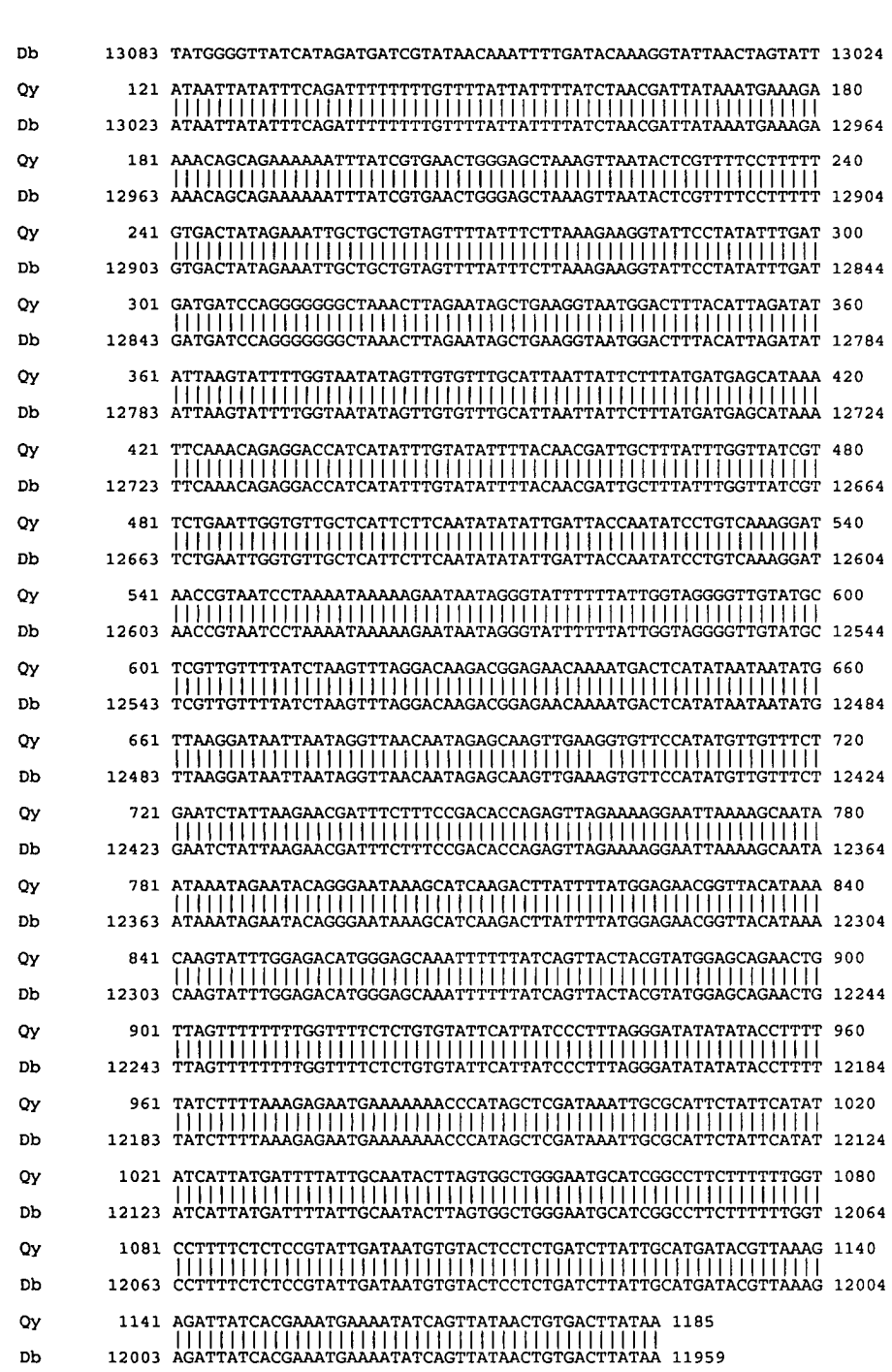
Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||||||||||||||||||||||||||||||||||||||
Db      12003 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 11959
```

```
RESULT 2
US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Patent No. 6855814
;   GENERAL INFORMATION:
;        APPLICANT: Blattner, Frederick R.
;                   Burland,     Valerie
;                   Perna,  Nicole T.
;                   Plunkett, Guy
;                   Welch,  Rod
;        TITLE OF INVENTION: No. 6855814el Sequences of E. coli O157
;        NUMBER OF SEQUENCES: 265
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Quarles & Brady
;             STREET: 1 South Pinckney Street
;             CITY: Madison
;             STATE: WI
;             COUNTRY: US
;             ZIP: 53701-2113
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Word Perfect 8.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/114,170
;             FILING DATE: 01-Apr-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/453,702
;             FILING DATE: 03-DEC-1999
;             APPLICATION NUMBER: 60/110,955
;             FILING DATE: 04-DEC-1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seay, Nicholas J.
;             REGISTRATION NUMBER: 27386
;             REFERENCE/DOCKET NUMBER: 960296.95017
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (608) 251-5000
;             TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 121:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 14187
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA  (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121
```


```
  Query Match             99.9%;  Score 1183.4;  DB 3;  Length 14187;
  Best Local Similarity   99.9%;  Pred. No. 1.9e-252;
  Matches 1184;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

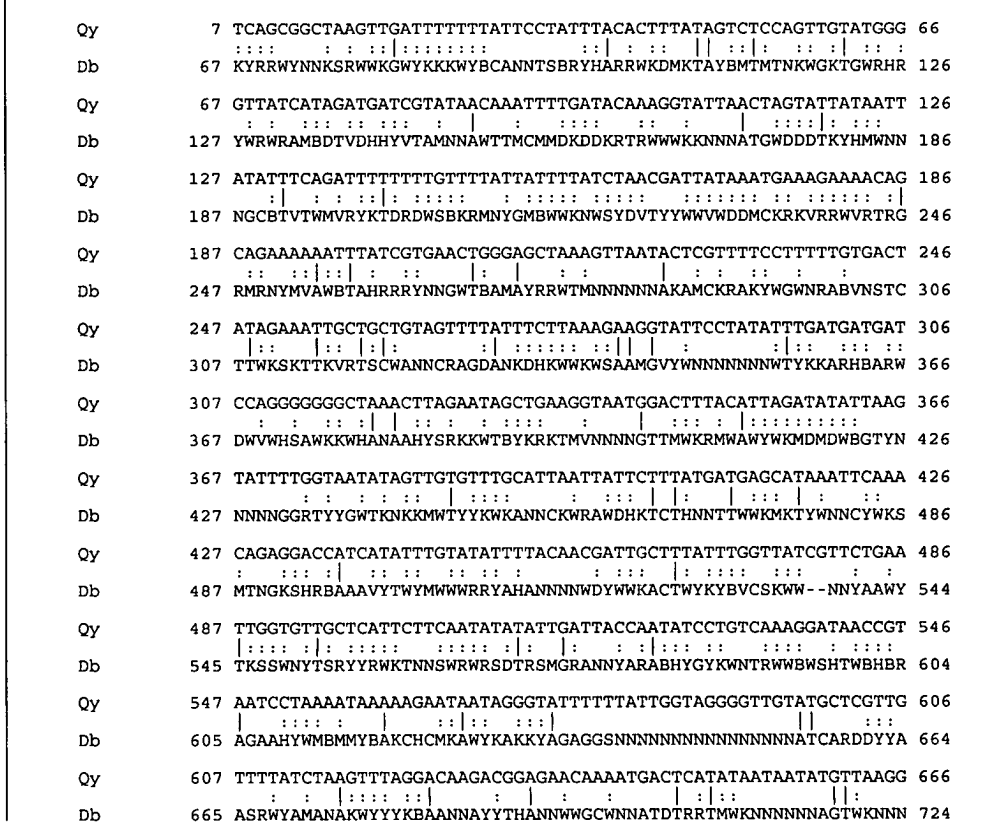
Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13143 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 13084

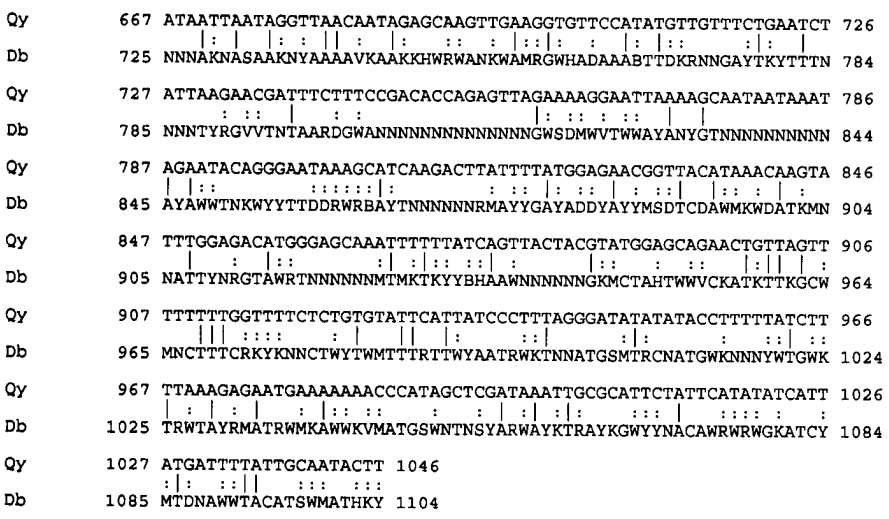
Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



RESULT 3
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

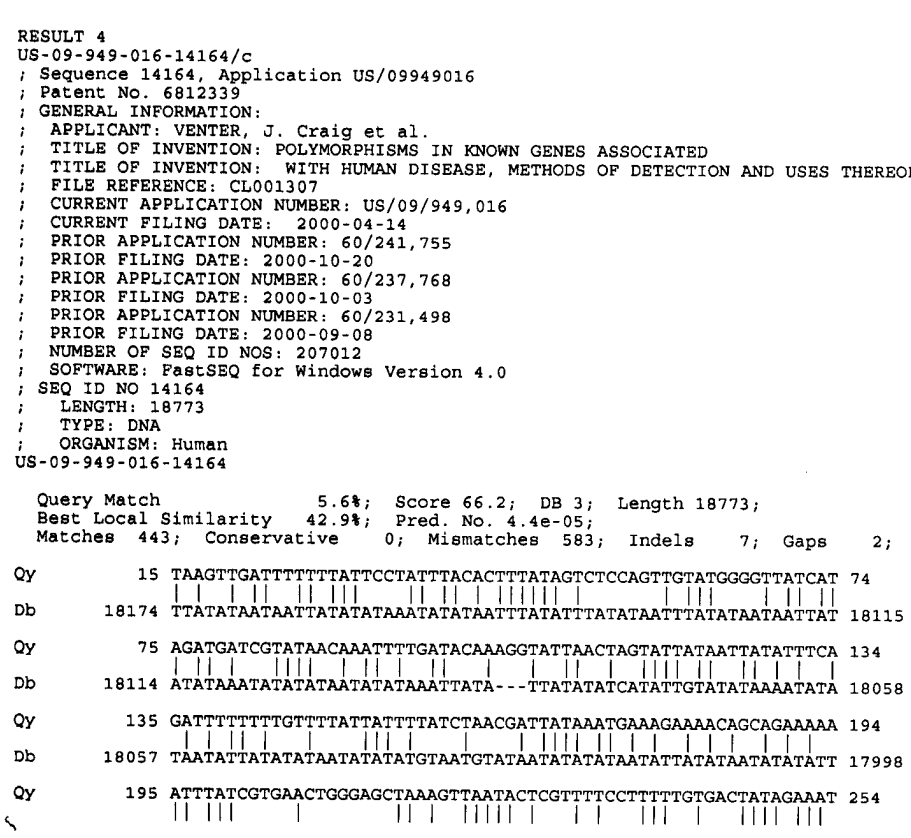
Query Match 6.5%; Score 77.4; DB 3; Length 1141;
Best Local Similarity 10.6%; Pred. No. 7.3e-08;
Matches 110; Conservative 392; Mismatches 536; Indels 2; Gaps 1;

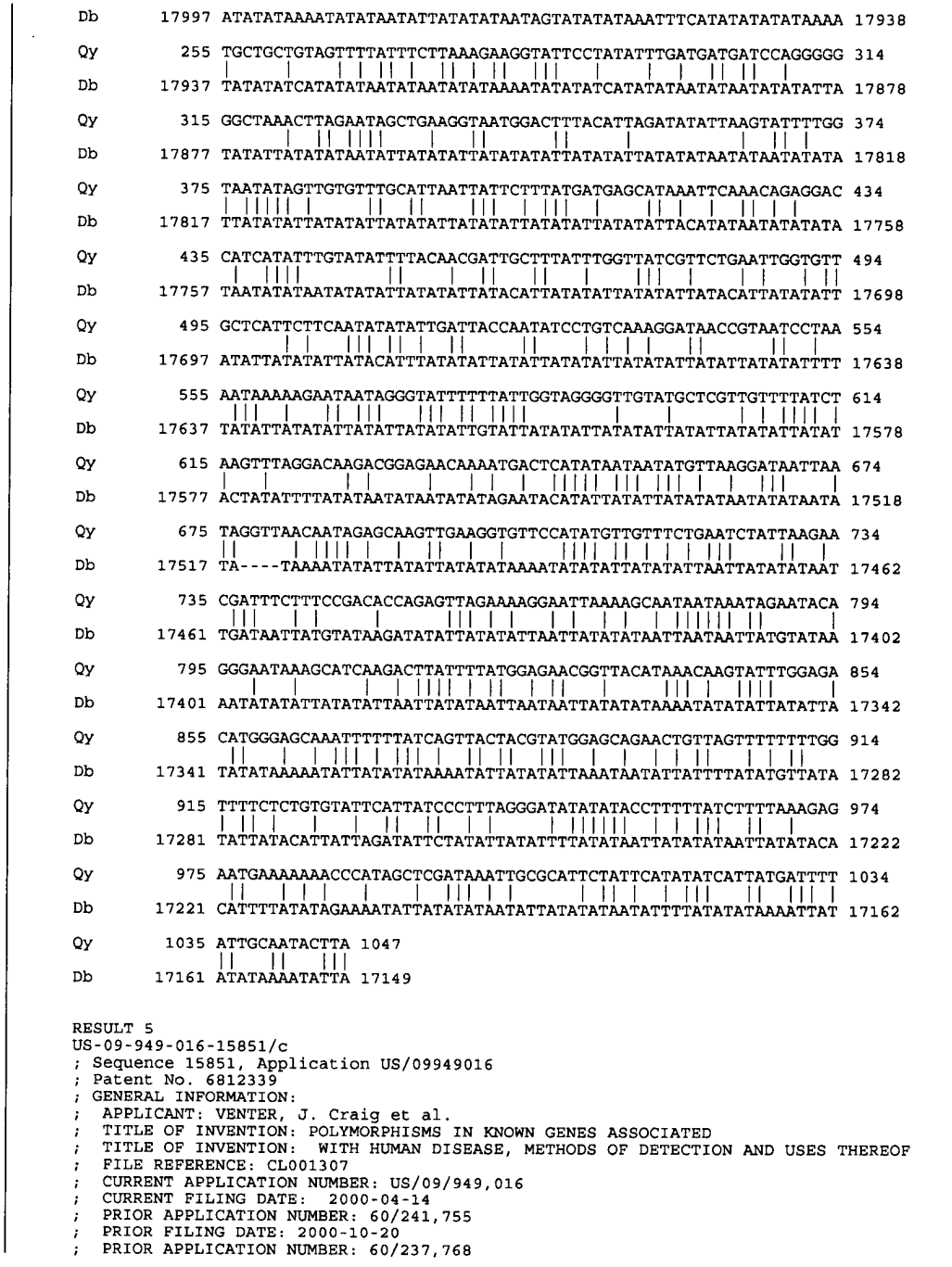




RESULT 4
US-09-949-016-14164/c
; Sequence 14164, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14164
; LENGTH: 18773
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14164

Query Match 5.6%; Score 66.2; DB 3; Length 18773;
Best Local Similarity 42.9%; Pred. No. 4.4e-05;
Matches 443; Conservative 0; Mismatches 583; Indels 7; Gaps 2;





RESULT 5
US-09-949-016-15851/c
; Sequence 15851, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768

```
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15851
;   LENGTH: 205044
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(205044)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15851

  Query Match             5.4%;  Score 64.2;  DB 3;  Length 205044;
  Best Local Similarity   44.9%;  Pred. No. 0.00022;
  Matches  294;  Conservative    0;  Mismatches  353;  Indels    8;  Gaps    1;

Qy         80 ATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTT 139
Db     201093 ATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATAT 201034

Qy        140 TTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTA 199
Db     201033 ATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAAT 200974

Qy        200 TCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTG 259
Db     200973 ATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAA 200914

Qy        260 CTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTA 319
Db     200913 ATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTA 200854

Qy        320 AACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATA 379
Db     200853 AAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATT 200794

Qy        380 TAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCA 439
Db     200793 TAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTA 200734

Qy        440 TATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTC- 498
Db     200733 TATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATAT 200674

Qy        499 -------ATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCC 551
Db     200673 AAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATT 200614

Qy        552 TAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTA 611
Db     200613 TAAAATATATATAAATATATTTTTATATATTTATATTTAAAATATATATAAATATATTTT 200554

Qy        612 TCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAAT 671
Db     200553 TATATATTTATATTTAAAATATATATAAATATATTTATATTTAAATATATTGATAAAATT 200494

Qy        672 TAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCT 726
Db     200493 TGATGATATAACCTAATATAATGTATAATAATTTAAATATATTTTGATTTATTTT 200439


RESULT 6
US-09-949-016-15852/c
; Sequence 15852, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15852
;   LENGTH: 205044
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(205044)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15852

  Query Match             5.4%;  Score 64.2;  DB 3;  Length 205044;
  Best Local Similarity   44.9%;  Pred. No. 0.00022;
  Matches  294;  Conservative    0;  Mismatches  353;  Indels    8;  Gaps    1;

Qy         80 ATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTT 139
Db     201093 ATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATAT 201034

Qy        140 TTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTA 199
Db     201033 ATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAAT 200974

Qy        200 TCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTG 259
Db     200973 ATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAA 200914

Qy        260 CTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTA 319
Db     200913 ATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTA 200854

Qy        320 AACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATA 379
Db     200853 AAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATT 200794

Qy        380 TAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCA 439
Db     200793 TAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTA 200734

Qy        440 TATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTC- 498
Db     200733 TATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATAT 200674

Qy        499 -------ATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCC 551
Db     200673 AAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATT 200614

Qy        552 TAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTA 611
Db     200613 TAAAATATATATAAATATATTTTTATATATTTATATTTAAAATATATATAAATATATTTT 200554

Qy        612 TCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAAT 671
Db     200553 TATATATTTATATTTAAAATATATATAAATATATTTATATTTAAATATATTGATAAAATT 200494

Qy        672 TAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCT 726
Db     200493 TGATGATATAACCTAATATAATGTATAATAATTTAAATATATTTTGATTTATTTT 200439


RESULT 7
US-09-949-016-15853/c
; Sequence 15853, Application US/09949016
; Patent No. 6812339
```

```
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15853
;   LENGTH: 205044
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(205044)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15853

  Query Match             5.4%;  Score 64.2;  DB 3;  Length 205044;
  Best Local Similarity   44.9%;  Pred. No. 0.00022;
  Matches  294;  Conservative    0;  Mismatches  353;  Indels    8;  Gaps    1;

Qy         80 ATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTT 139
Db     201093 ATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATAT 201034

Qy        140 TTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTA 199
Db     201033 ATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAAT 200974

Qy        200 TCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTG 259
Db     200973 ATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAA 200914

Qy        260 CTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTA 319
Db     200913 ATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTA 200854

Qy        320 AACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATA 379
Db     200853 AAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATT 200794

Qy        380 TAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCA 439
Db     200793 TAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTA 200734

Qy        440 TATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTC- 498
Db     200733 TATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATAT 200674

Qy        499 -------ATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCC 551
Db     200673 AAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATT 200614

Qy        552 TAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTA 611
Db     200613 TAAAATATATATAAATATATTTTTATATATTTATATTTAAAATATATATAAATATATTTT 200554

Qy        612 TCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAAT 671
Db     200553 TATATATTTATATTTAAAATATATATAAATATATTTATATTTAAATATATTGATAAAATT 200494

Qy        672 TAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCT 726
Db     200493 TGATGATATAACCTAATATAATGTATAATAATTTAAATATATTTTGATTTATTTT 200439
```

```
RESULT 8
US-09-949-016-12387/c
; Sequence 12387, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12387
;   LENGTH: 223471
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(223471)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12387

  Query Match             5.4%;  Score 64.2;  DB 3;  Length 223471;
  Best Local Similarity   44.9%;  Pred. No. 0.00023;
  Matches  294;  Conservative    0;  Mismatches  353;  Indels    8;  Gaps    1;

Qy         80 ATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTT 139
Db     171520 ATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATAT 171461

Qy        140 TTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTA 199
Db     171460 ATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAAT 171401

Qy        200 TCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTG 259
Db     171400 ATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAA 171341

Qy        260 CTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTA 319
Db     171340 ATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTA 171281

Qy        320 AACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATA 379
Db     171280 AAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATT 171221

Qy        380 TAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCA 439
Db     171220 TAAAATATATATAAATATATTTTTATATATTTATATTTAAAATATATATAAATATATTTTA 171161

Qy        440 TATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTC- 498
Db     171160 TATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATAT 171101

Qy        499 -------ATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCC 551
Db     171100 AAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATT 171041

Qy        552 TAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTA 611
Db     171040 TAAAATATATATAAATATATTTTTATATATTTATATTTAAAATATATATAAATATATTTT 170981

Qy        612 TCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAAT 671
Db     170980 TATATATTTATATTTAAAATATATATAAATATATTTATATTTAAATATATTGATAAAATT 170921
```

```
Qy        672 TAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCT 726
Db     170920 TGATGATATAACCTAATATAATGTATAATAATTTAAATATATTTTGATTTATTTT 170866
```

RESULT 9
US-09-949-016-12724/c
; Sequence 12724, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12724
; LENGTH: 223471
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(223471)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12724

Query Match 5.4%; Score 64.2; DB 3; Length 223471;
Best Local Similarity 44.9%; Pred. No. 0.00023;
Matches 294; Conservative 0; Mismatches 353; Indels 8; Gaps 1;

```
Qy         80 ATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTT 139
Db     171520 ATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATAT 171461

Qy        140 TTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTA 199
Db     171460 ATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAAT 171401

Qy        200 TCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTG 259
Db     171400 ATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAA 171341

Qy        260 CTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTA 319
Db     171340 ATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTA 171281

Qy        320 AACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATA 379
Db     171280 AAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATT 171221

Qy        380 TAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCA 439
Db     171220 TAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTA 171161

Qy        440 TATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTC- 498
Db     171160 TATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATAT 171101

Qy        499 -------ATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCC 551
Db     171100 AAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATT 171041

Qy        552 TAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTA 611
```

```
Db     171040 TAAAATATATATAAATATATTTTTATATATTTATATTTAAAATATATATAAATATATTTT 170981

Qy        612 TCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAAT 671
Db     170980 TATATATTTATATTTAAAATATATATAAATATATTTATATTTAAATATATTGATAAAATT 170921

Qy        672 TAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCT 726
Db     170920 TGATGATATAACCTAATATAATGTATAATAATTTAAATATATTTTGATTTATTTT 170866
```

RESULT 10
US-09-949-016-12725/c
; Sequence 12725, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12725
; LENGTH: 223471
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(223471)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12725

Query Match 5.4%; Score 64.2; DB 3; Length 223471;
Best Local Similarity 44.9%; Pred. No. 0.00023;
Matches 294; Conservative 0; Mismatches 353; Indels 8; Gaps 1;

```
Qy         80 ATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTT 139
Db     171520 ATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATAT 171461

Qy        140 TTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTA 199
Db     171460 ATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAAT 171401

Qy        200 TCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTG 259
Db     171400 ATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAA 171341

Qy        260 CTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTA 319
Db     171340 ATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTA 171281

Qy        320 AACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATA 379
Db     171280 AAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATT 171221

Qy        380 TAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCA 439
Db     171220 TAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTA 171161

Qy        440 TATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTC- 498
Db     171160 TATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATAT 171101

Qy        499 -------ATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCC 551
```

```
Db     171100 AAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATT 171041

Qy        552 TAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTA 611

Db     171040 TAAAATATATATAAATATATTTTTATATATTTATATTTAAAATATATATAAATATATTTT 170981

Qy        612 TCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAAT 671

Db     170980 TATATATTTATATTTAAAATATATATAAATATATTTATATTTAAATATATTGATAAAATT 170921

Qy        672 TAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCT 726

Db     170920 TGATGATATAACCTAATATAATATGTATAATAATTTAAATATATTTTGATTTATTTT 170866


RESULT 11
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             4.9%;  Score 58.6;  DB 3;  Length 1141;
  Best Local Similarity   10.1%;  Pred. No. 0.0011;
  Matches  106;  Conservative  373;  Mismatches  574;  Indels    0;  Gaps    0;

Qy         19 TTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATAGAT 78

Db       1129 TTKTKTYKKANNNNNNNNGMGKDWNRMDATKWSATGTAWWTNHAKRGATMCWYWYWTGTN 1070

Qy         79 GATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATT 138

Db       1069 RRWCMRTYAMRTWYTRSNANWSCATKBMWWTMKWYATKYRTAWYAMWCAWRNNNMWCATN 1010

Qy        139 TTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTT 198

Db       1009 GYAKSCATNNAMWYATTRWAAYAAAKWARWAGNNMRMYGAAAGNKWGCMAAMATMGBWWA 950

Qy        199 ATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCT 258

Db        949 DTAGKMCNNNNNNWTTDVRRMAMKAKNNNNNNAYWTACYNRAATNNKMATHWMKWTHGAH 890

Qy        259 GCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCT 318

Db        889 SKRRTRHHTRTCRRTKYNNNNNNARTVYWYHHAARRWMNAWWTRTNNNNNNNNNNACRNT 830

Qy        319 AAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAAT 378

Db        829 RTWWABWKHSWCNNNNNNNNNNNNNNNTWCHYTTANABBCYRANNNNAAARMARTCNNYMH 770

Qy        379 ATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATC 438

Db        769 AAVTTTHTDWCYKTWMNTWYWDMMTTMBTTTTRNMTTSTNMTNNNNNNMWACTNNNNNNM 710

Qy        439 ATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTC 498

Db        709 WKAYYAHATNNWGCWWNNTDARRTNNTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNNNN 650

Qy        499 ATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATA 558

Db        649 NNNNNNNNNNNSCCTCTRMMTMRWTMKGDGMTVRKKVKWRDTTCTYVDVWADSWVWWYAN 590

Qy        559 AAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGT 618

Db        589 WMRCRDVTYTRNNTYCKSYAHSYWYWSNNAMWYRRYSARNWSSMARWTTRNNWWMSGBVR 530

Qy        619 TTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAATTAATAGG 678

Db        529 MRWAGTMWWRHWNNNNTDTRYYWWWKRWARBTTTVYDSMCNAKSMWRGNNWRAMKMWWAA 470

Qy        679 TTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCTATTAAGAACGAT 738

Db        469 NNDAGAMDHWTYWMGNNTMWMRRAWKMMNMAWCRRAYCCNNNNNRACVWHKHKMWRWTWK 410

Qy        739 TTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATAATAAATAGAATACAGGGA 798

Db        409 YMWKAACNNNNBKAMYMRVAWMMYSRDTTNTDWMMWTSDWBWHWYTVDYTMMRAWNNNNN 350

Qy        799 ATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAACAAGTATTTGGAGACATG 858

Db        349 NNWRBCKTTSWMWWMDHMNTHCTYGNNTWGSAYBMAAMSMWAAGASNBVTYNWCWRMTYM 290

Qy        859 GGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTGTTAGTTTTTTTTGGTTTT 918

Db        289 GKTMTNNNNNNKAWYYRTKTVAWCNNRYYYDTAVWTBKRNYKYCYAYBWYYBMYMGKHHW 230

Qy        919 CTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTTTATCTTTTAAAGAGAATG 978

Db        229 BWWRRABHRSWNMWWVKCRNKYMVSWHYHAMRYBKWABAVGCNNNWKDRMAHHHWCATNN 170

Qy        979 AAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATATATCATTATGATTTTATTG 1038

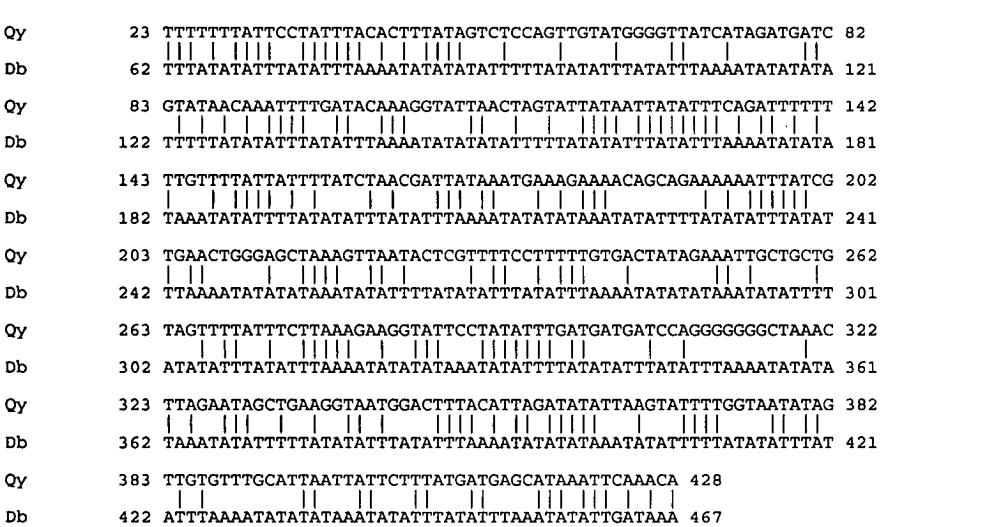
Db        169 NMMWWWYAYMHHMHKKGKAAWTNNKTABRDDHBAHVKTYWYWRYDYWCAMCWMNAKAKVR 110

Qy       1039 CAATACTTAGTGGCTGGGAATGCATCGGCCTTC 1071

Db        109 TAMKHMWYYTDRYVSANNTGVRWMMMRWCMWWY 77

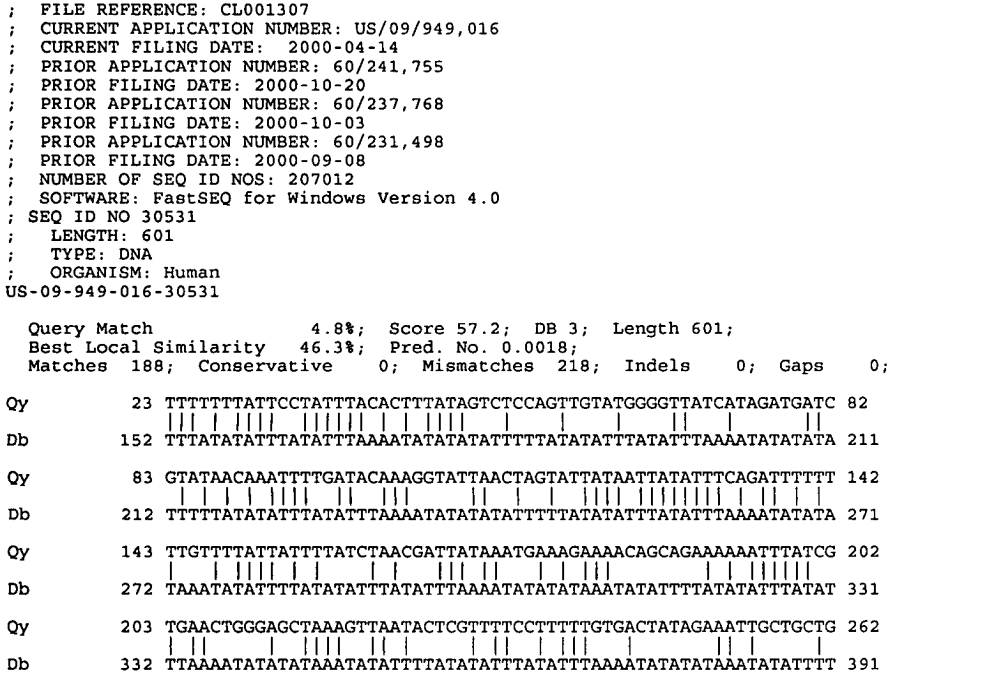

RESULT 12
US-09-949-016-30530
; Sequence 30530, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30530
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-30530

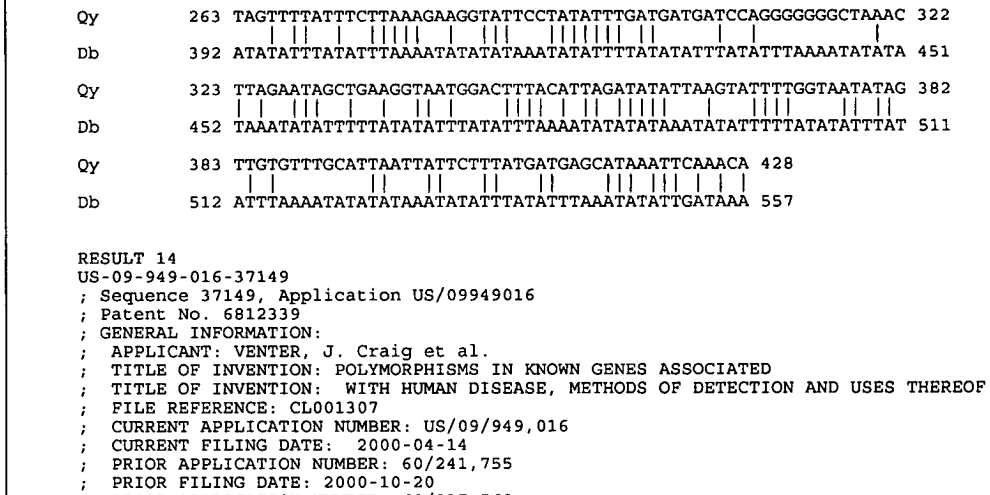
  Query Match             4.8%;  Score 57.2;  DB 3;  Length 601;
  Best Local Similarity   46.3%;  Pred. No. 0.0018;
  Matches  188;  Conservative    0;  Mismatches  218;  Indels    0;  Gaps    0;
```



RESULT 13
US-09-949-016-30531
; Sequence 30531, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30531
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-30531

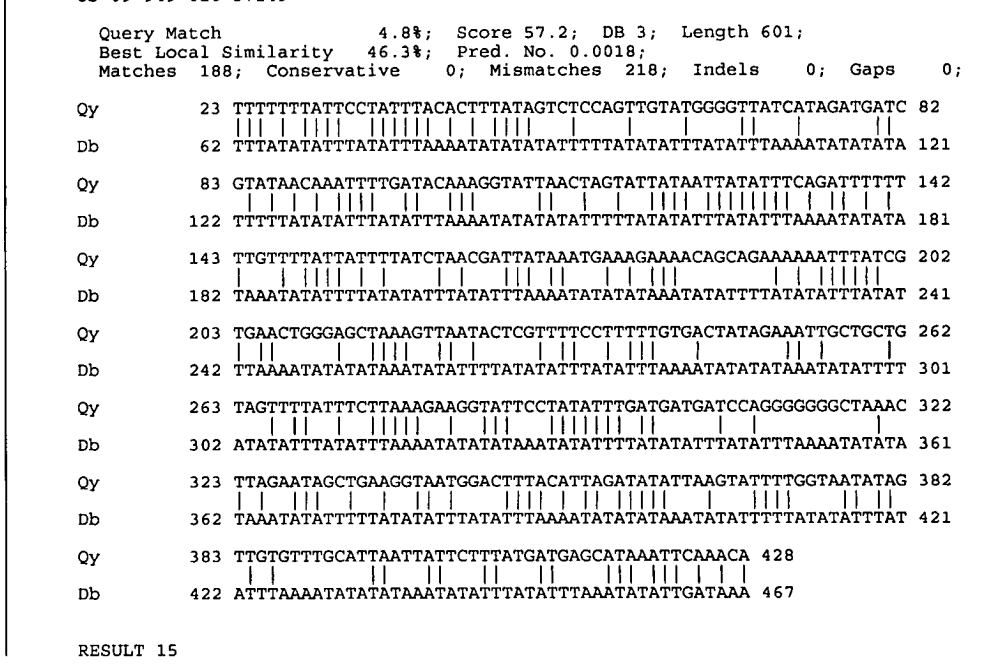
Query Match 4.8%; Score 57.2; DB 3; Length 601;
Best Local Similarity 46.3%; Pred. No. 0.0018;
Matches 188; Conservative 0; Mismatches 218; Indels 0; Gaps 0;





RESULT 14
US-09-949-016-37149
; Sequence 37149, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37149
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-37149

Query Match 4.8%; Score 57.2; DB 3; Length 601;
Best Local Similarity 46.3%; Pred. No. 0.0018;
Matches 188; Conservative 0; Mismatches 218; Indels 0; Gaps 0;



RESULT 15

```
US-09-949-016-37150
; Sequence 37150, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37150
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-37150

  Query Match             4.8%;  Score 57.2;  DB 3;  Length 601;
  Best Local Similarity   46.3%;  Pred. No. 0.0018;
  Matches  188;  Conservative    0;  Mismatches  218;  Indels    0;  Gaps    0;

Qy         23 TTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATAGATGATC 82
              ||| | ||||  |||||| | | ||||   |    |    |    ||   |      ||
Db        152 TTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATA 211

Qy         83 GTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTTTTT 142
              | | | | ||||  ||  |||     ||  |  |  |||| |||||||| | || | |
Db        212 TTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATA 271

Qy        143 TTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTATCG 202
              |   | |||| | |   | |   ||| ||   | | |||        | | ||||||
Db        272 TAAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATAT 331

Qy        203 TGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTGCTG 262
              | ||      | ||||  || |      | ||  | |||   |        || |      |
Db        332 TTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTT 391

Qy        263 TAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTAAAC 322
              | ||  |  |||||  |  |||   |||||| ||      |  |           |
Db        392 ATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATA 451

Qy        323 TTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATATAG 382
              | |  |||  |  |  || |   |||| | || |||||   |  ||||    || ||
Db        452 TAAATATATTTTTATATATTTATATTTAAAATATATATAAATATATTTTTATATATTTAT 511

Qy        383 TTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACA 428
              | |       ||   ||   ||   ||    ||| ||| | | |
Db        512 ATTTAAAATATATATAAATATATTTATATTTAAATATATTGATAAA 557


Search completed: May 11, 2006, 05:23:08
Job time : 169.291 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:01:11 ; Search time 782.135 Seconds
(without alignments)
12528.801 Million cell updates/sec

Title: US-09-423-093B-2_COPY_858_2042
Perfect score: 1185
Sequence: 1 gtgaagtcagcggctaagtt.....gttataactgtgacttataa 1185

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1185	100.0	14024	3	US-09-423-093A-2	Sequence 2, Appli
c 2	1183.4	99.9	14187	5	US-10-114-170-121	Sequence 121, App
3	75.2	6.3	115218	8	US-10-278-698-255	Sequence 255, App
4	75.2	6.3	115218	8	US-10-278-698-769	Sequence 769, App
c 5	74.8	6.3	115218	8	US-10-278-698-255	Sequence 255, App
c 6	74.8	6.3	115218	8	US-10-278-698-769	Sequence 769, App
7	71	6.0	8056	8	US-10-473-126-240	Sequence 240, App
8	70	5.9	8056	8	US-10-473-126-386	Sequence 386, App
9	69.2	5.8	3673778	6	US-10-312-841-1	Sequence 1, Appli
c 10	68.6	5.8	6816	8	US-10-723-860-7676	Sequence 7676, Ap
11	68.4	5.8	5728	6	US-10-311-455-74	Sequence 74, Appl
c 12	68	5.7	38678	9	US-10-893-315-136	Sequence 136, App
c 13	68	5.7	38684	9	US-10-893-315-154	Sequence 154, App
14	68	5.7	3673778	6	US-10-312-841-2	Sequence 2, Appli
15	67.8	5.7	14551	6	US-10-240-485-138	Sequence 138, App
16	67.6	5.7	8222	9	US-10-486-319A-65	Sequence 65, Appl
17	67	5.7	15500	7	US-10-221-714A-230	Sequence 230, App
18	66.4	5.6	6255	6	US-10-311-455-933	Sequence 933, App
19	66.4	5.6	15732	5	US-10-239-676-95	Sequence 95, Appl
20	66.4	5.6	15732	6	US-10-240-453-107	Sequence 107, App
21	66.2	5.6	6954	6	US-10-311-455-1363	Sequence 1363, Ap
22	66	5.6	8222	9	US-10-486-319A-43	Sequence 43, Appl
23	65	5.5	15548	6	US-10-311-455-2128	Sequence 2128, Ap
24	63.8	5.4	13131	7	US-10-240-589C-58	Sequence 58, Appl
25	63.4	5.4	18683	6	US-10-311-455-286	Sequence 286, App
26	63.4	5.4	18683	6	US-10-240-452-34	Sequence 34, Appl
27	63.4	5.4	46951	6	US-10-091-281-2	Sequence 2, Appli
28	62.4	5.3	9510	6	US-10-240-485-115	Sequence 115, App
29	62.4	5.3	9510	7	US-10-221-714A-159	Sequence 159, App
30	62.2	5.2	14924	6	US-10-311-455-197	Sequence 197, App
31	62.2	5.2	14924	6	US-10-240-452-21	Sequence 21, Appl
32	61.8	5.2	5930	6	US-10-311-455-490	Sequence 490, App
33	61.6	5.2	6534	5	US-10-239-676-111	Sequence 111, App
34	61.6	5.2	6534	6	US-10-311-455-1399	Sequence 1399, Ap
35	61.6	5.2	6534	6	US-10-240-453-125	Sequence 125, App
36	61.2	5.2	6641	6	US-10-311-455-288	Sequence 288, App
37	61.2	5.2	6641	6	US-10-240-452-36	Sequence 36, Appl
38	61.2	5.2	7104	6	US-10-240-485-142	Sequence 142, App
39	60.6	5.1	15399	6	US-10-311-455-1488	Sequence 1488, Ap
40	60.4	5.1	8201	6	US-10-311-455-279	Sequence 279, App
41	60.4	5.1	8201	6	US-10-240-452-27	Sequence 27, Appl
42	60	5.1	3089	9	US-10-954-778-160	Sequence 160, App
43	60	5.1	4985	5	US-10-094-240-10	Sequence 10, Appl
44	60	5.1	4985	6	US-10-056-405-10	Sequence 10, Appl
45	60	5.1	16545	6	US-10-311-455-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-09-423-093A-2
; Sequence 2, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 14024
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-423-093A-2

Query Match 100.0%; Score 1185; DB 3; Length 14024;
Best Local Similarity 100.0%; Pred. No. 1.6e-216;
Matches 1185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 917

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 977

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 1037

Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 1097

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 1157
```

```
Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 1217

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 1277

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 1337

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 1397

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 1457

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 1517

Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 1577

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1578 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 1637

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1638 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAA 1697

Qy        841 CAAGTATTTGGAGACATGGGAGCAAATTTTTTTATCAGTTACTACGTATGGAGCAGAACTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1698 CAAGTATTTGGAGACATGGGAGCAAATTTTTTTATCAGTTACTACGTATGGAGCAGAACTG 1757

Qy        901 TTAGTTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1758 TTAGTTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTT 1817

Qy        961 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1818 TATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATAT 1877

Qy       1021 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1878 ATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCATCGGCCTTCTTTTTTGGT 1937

Qy       1081 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1938 CCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTATTGCATGATACGTTAAAG 1997

Qy       1141 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 1185
              ||||||||||||||||||||||||||||||||||||||||||||
Db       1998 AGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTATAA 2042
```

RESULT 2
US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14187
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121

Query Match 99.9%; Score 1183.4; DB 5; Length 14187;
Best Local Similarity 99.9%; Pred. No. 3.2e-216;
Matches 1184; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13143 GTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTG 13084

Qy         61 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13083 TATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATT 13024

Qy        121 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13023 ATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGA 12964

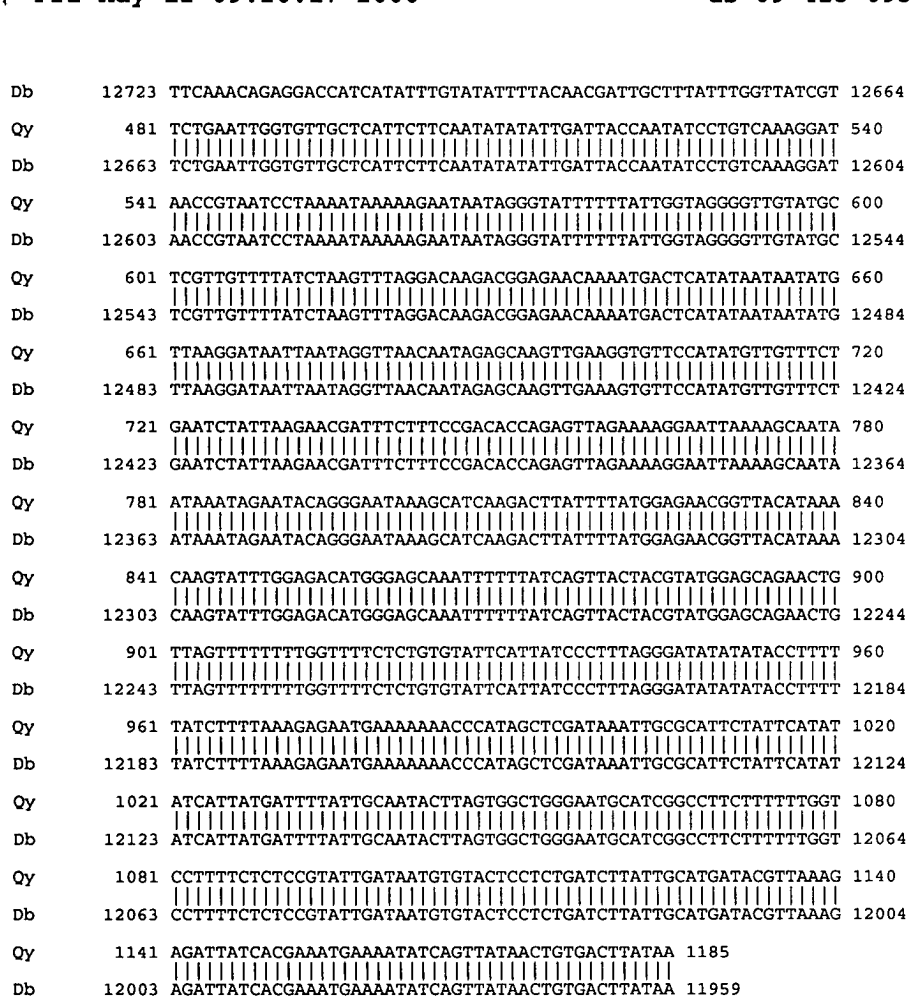
Qy        181 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12963 AAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTT 12904

Qy        241 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12903 GTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGAT 12844

Qy        301 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12843 GATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATAT 12784

Qy        361 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12783 ATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 12724

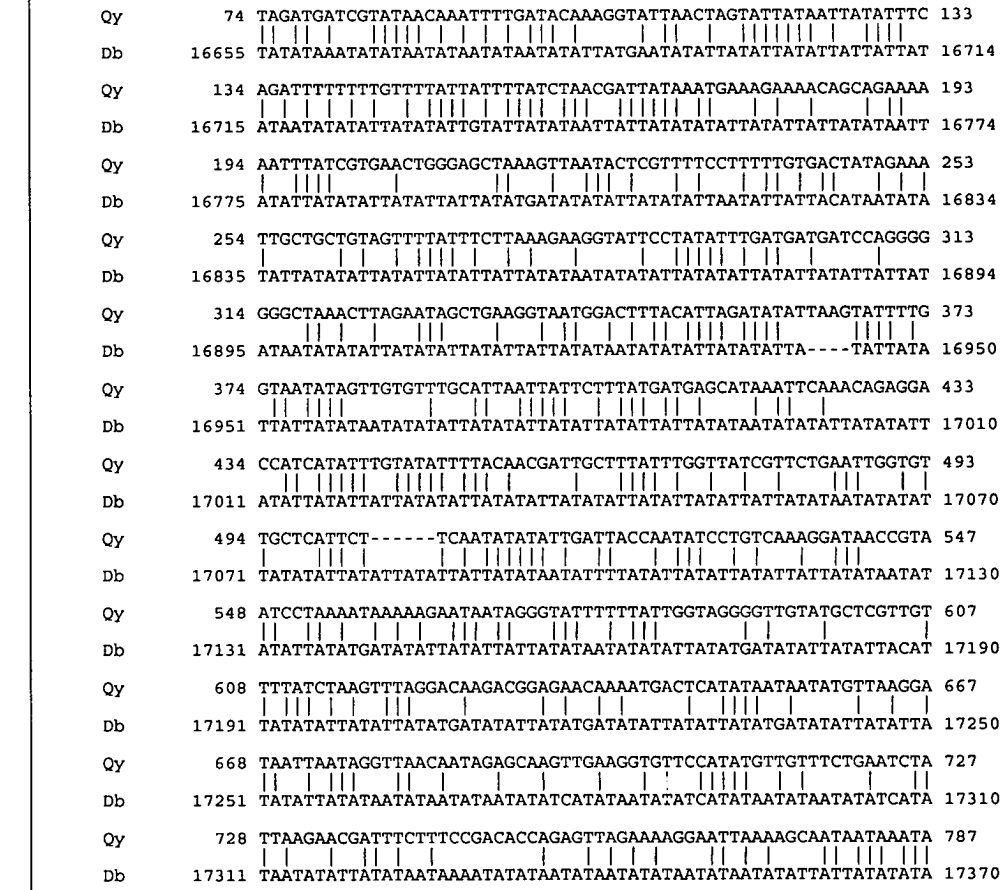
Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



RESULT 3
US-10-278-698-255
; Sequence 255, Application US/10278698
; Publication No. US20050037344A1
; GENERAL INFORMATION:
; APPLICANT: PathoArray GmbH
; APPLICANT: Stuhlmuller, Bruno
; APPLICANT: Haupl, Thomas
; TITLE OF INVENTION: Nucleic Acid Array
; FILE REFERENCE: O30027US
; CURRENT APPLICATION NUMBER: US/10/278,698
; CURRENT FILING DATE: 2002-10-23
; NUMBER OF SEQ ID NOS: 1050
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 255
; LENGTH: 115218
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-278-698-255

Query Match 6.3%; Score 75.2; DB 8; Length 115218;
Best Local Similarity 43.9%; Pred. No. 0.0011;

Matches 428; Conservative 0; Mismatches 538; Indels 10; Gaps 2;



RESULT 4
US-10-278-698-769

; Sequence 769, Application US/10278698
; Publication No. US20050037344A1
; GENERAL INFORMATION:
; APPLICANT: PathoArray GmbH
; APPLICANT: Stuhlmuller, Bruno
; APPLICANT: Haupl, Thomas
; TITLE OF INVENTION: Nucleic Acid Array
; FILE REFERENCE: O30027US
; CURRENT APPLICATION NUMBER: US/10/278,698
; CURRENT FILING DATE: 2002-10-23
; NUMBER OF SEQ ID NOS: 1050
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 769
; LENGTH: 115218
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-278-698-769

Query Match 6.3%; Score 75.2; DB 8; Length 115218;
Best Local Similarity 43.9%; Pred. No. 0.0011;
Matches 428; Conservative 0; Mismatches 538; Indels 10; Gaps 2;

```
Qy        74 TAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTC 133
             || || |  |||||| | | | | ||| |     | ||  |  |||||||  |  ||||
Db     16655 TATATAAATATATAATATAATATAATATATTATGAATATATTATATTATATTATTATTAT 16714

Qy       134 AGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAA 193
             | | | | | |  | |||| | |||| |||  |||||| ||   |  |   |    | ||
Db     16715 ATAATATATATTATATATTGTATTATATAATTATTATATATATTATATTATTATATAATT 16774

Qy       194 AATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAA 253
             |  ||||      |         ||   |  ||| |   | |   | || | ||   | | |
Db     16775 ATATTATATATTATATTATTATATGATATATATTATATATTAATATTATTACATAATATA 16834

Qy       254 TTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGG 313
             |      | |  | |||| |  | |    |     | ||||| | ||  |     |
Db     16835 TATTATATATTATATTATATTATTATATAATATATATTATATATTATATTATATTATTAT 16894

Qy       314 GGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTG 373
               || |  |  |||     |   | ||  | | || |||| ||||      |||| |
Db     16895 ATAATATATATTATATATTATATTATATAATATATATTATATATTA----TATTATA 16950

Qy       374 GTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGA 433
             || ||||      |   ||  |||||  | ||| || |    | ||  |   |
Db     16951 TTATTATATAATATATATTATATATTATATTATTATTATATAATATATTATATATT 17010

Qy       434 CCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGT 493
             || |||||  |||||  ||| |      |  |||| |  |  |  |    |||   | |
Db     17011 ATATTATATTATTATATATTATATATTATATATTATATTATTATATAATATATAT 17070

Qy       494 TGCTCATTCT------TCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTA 547
             |    ||| |      | | |||||| |  ||   | ||| | |   |    |
Db     17071 TATATATTATATTATATTATTATATAATATTTTATATTATATTATTATTATATAATAT 17130

Qy       548 ATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGT 607
             ||  || |  | | |  ||| ||  |||  | |||  |  |  |  |  |
Db     17131 ATATTATATGATATATTATATTATTATATAATATATATTATATGATATATTATATTACAT 17190

Qy       608 TTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGA 667
             | ||| | |  |||    |      | |  | |     | ||||  |     |  | |
Db     17191 TATATATTATATTATATGATATATTATATGATATATTATATTATATGATATATTATATTA 17250

Qy       668 TAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCTA 727
             ||  | |||    ||  |    |   |  |  |  |  |||||  | |    |   ||
Db     17251 TATATTATATAATATAATATAATATATCATATAATATATCATATAATATAATATATCATA 17310

Qy       728 TTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATAATAAATA 787
             | |   | || |  |          |  | | | |  || | |  || ||| |||
Db     17311 TAATATATTATATAATAAAATATATAATATAATATATAATATAATATATTATTATATATA 17370

Qy       788 GAATACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAACAAGTAT 847
             ||||  |   ||||   || |    |  | ||  | ||   |   | | |
Db     17371 TAATATATTATATAACATATTATATATATAATAATATATAATATTATATAATATATTATA 17430

```

```
Qy       848 TTGGAGACATGGGAGCAAATTTTTTATCAGTTACTACGTATGGAGCAGAACTGTTAGTTT 907
             ||  | | |    |  | | | | ||| |  |  ||          ||    | |
Db     17431 TTATATATAATATAATATATATTATATTATATTATATATTATTAATATATTATAA 17490

Qy       908 TTTTTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTTTATCTTT 967
             || |    | | | | |    | |||||     |       ||||||    || ||| | |
Db     17491 TTATATAATATATATTATTATATAATTATAATATATTAATAATATATCGTATTATATATAT 17550

Qy       968 TAAAGAGAATGAAAAAAACCCATAGCTCGATAAATTGCGCATTCTATTCATATATCATTA 1027
             || |   || | | |      ||    |||  |   ||| ||   ||||  ||||
Db     17551 TATATTATATTATATATTATATTATTAAAATATTATCTATATTATAAATATATTATATTA 17610

Qy      1028 TGATTTTATTGCAATA 1043
             |  |||||||  | ||
Db     17611 TATTTTTATTATATTA 17626
```

RESULT 5
US-10-278-698-255/c
; Sequence 255, Application US/10278698
; Publication No. US20050037344A1
; GENERAL INFORMATION:
; APPLICANT: PathoArray GmbH
; APPLICANT: Stuhlmuller, Bruno
; APPLICANT: Haupl, Thomas
; TITLE OF INVENTION: Nucleic Acid Array
; FILE REFERENCE: O30027US
; CURRENT APPLICATION NUMBER: US/10/278,698
; CURRENT FILING DATE: 2002-10-23
; NUMBER OF SEQ ID NOS: 1050
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 255
; LENGTH: 115218
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-278-698-255

Query Match 6.3%; Score 74.8; DB 8; Length 115218;
Best Local Similarity 43.6%; Pred. No. 0.0013;
Matches 428; Conservative 0; Mismatches 552; Indels 2; Gaps 2;

```
Qy        69 TATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTAT 128
             ||| ||| ||           | | | | |  ||  ||  || ||  ||| |||
Db     17607 TATAATATATTTATAATATAGATAATATTTTAATAATATAATATATAATATAATATAATA 17548

Qy       129 ATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCA 188
               |  |      | |  | ||| ||| ||||      || ||| ||  |  | |  | |
Db     17547 TATATAATACGATATATTATTAATATATTATAATTATATAATATATATATTATATAATTA 17488

Qy       189 GAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTAT 248
             || | |||  ||  |  |          ||  | |  |||  |   ||   || |    | |||
Db     17487 TAATATATTAATAATATATAATATAATATATATTATATTATATTATATATAATAT 17428

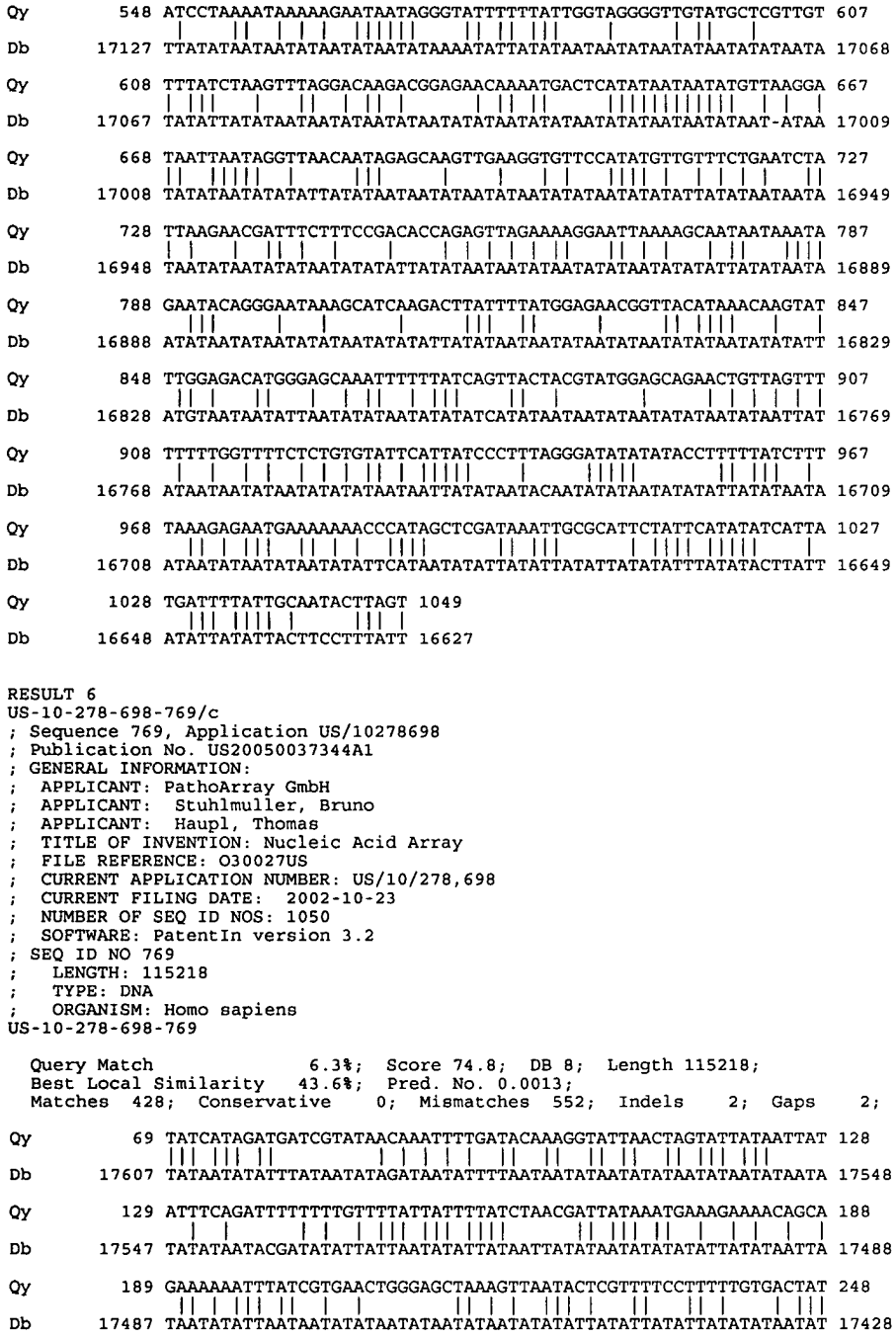
Qy       249 AGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCC 308
             |  |  |  |     || | |   || ||| | |    ||   |     |  || ||
Db     17427 AATATATTATATAATATTATATATTATTATATATATAATATGTTATATAATATATTATAT 17368

Qy       309 AGGGGGGGCTAAACT-TAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGT 367
             |          || | | ||  |||  | |   ||      ||   ||  |||||||  |
Db     17367 ATATAATAATATATTATATTATATATTATATTATATATTTTATTATATAATATATTATAT 17308

Qy       368 ATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAAC 427
               | |  ||   ||  ||   |  ||     |||  | |||  ||        |||  | |
Db     17307 GATATATTATATTATATGATATATTATATGATATATTATATTATATTATATAATATATAA 17248

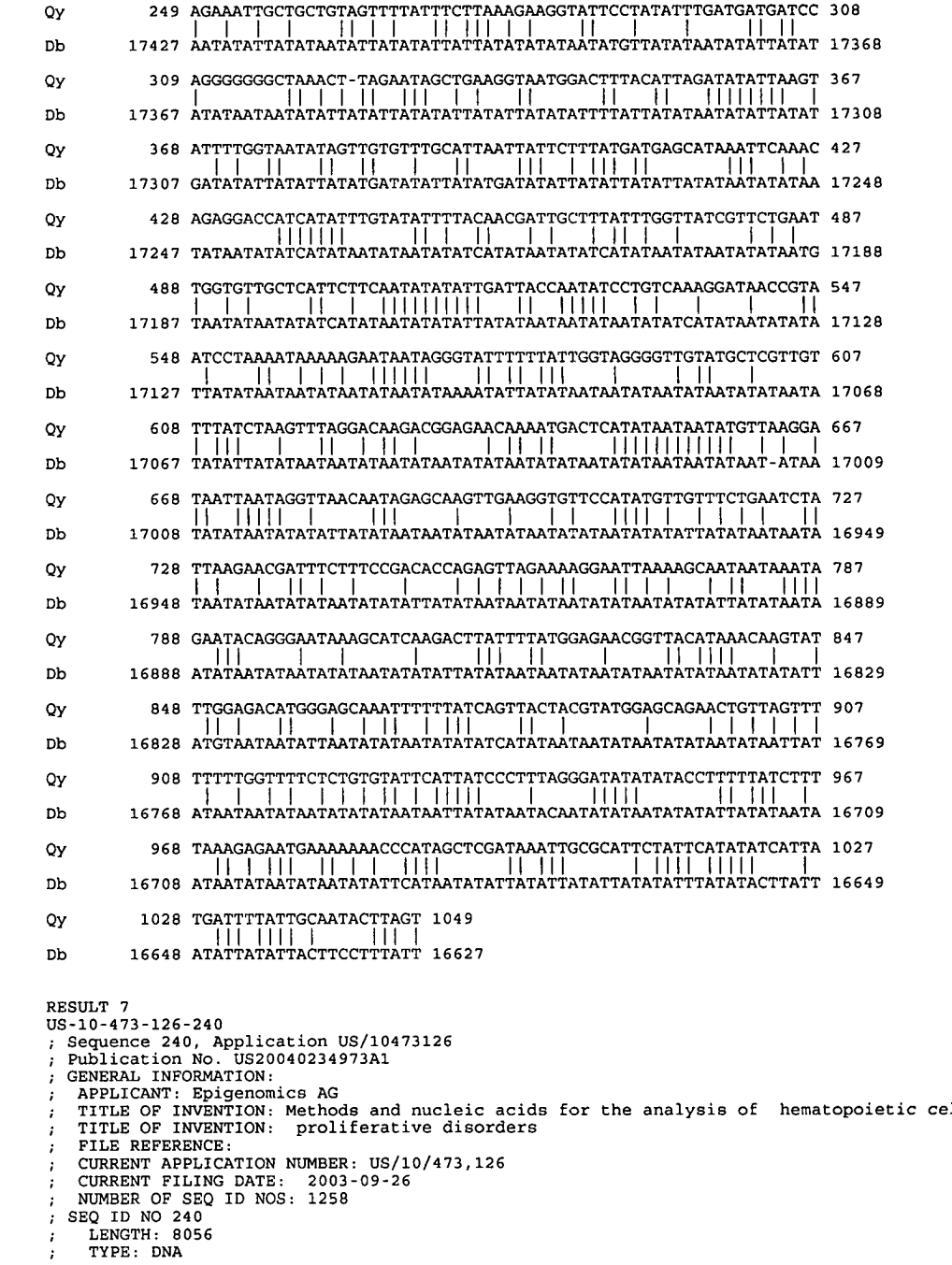
Qy       428 AGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAAT 487
             |||||||          || |  ||  | |  | || |  |  |  | |  | | |
Db     17247 TATAATATATCATATAATATAATATATCATATAATATATCATATAATATAATATATAATG 17188

Qy       488 TGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTA 547
             |  | |      ||  |  ||||||||||   ||  |||||  | |  |    |  ||
Db     17187 TAATATAATATATCATATAATATATATTATATAATAATATAATATATCATATAATATATA 17128
```



RESULT 6
US-10-278-698-769/c
; Sequence 769, Application US/10278698
; Publication No. US20050037344A1
; GENERAL INFORMATION:
; APPLICANT: PathoArray GmbH
; APPLICANT: Stuhlmuller, Bruno
; APPLICANT: Haupl, Thomas
; TITLE OF INVENTION: Nucleic Acid Array
; FILE REFERENCE: O30027US
; CURRENT APPLICATION NUMBER: US/10/278,698
; CURRENT FILING DATE: 2002-10-23
; NUMBER OF SEQ ID NOS: 1050
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 769
; LENGTH: 115218
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-278-698-769

Query Match 6.3%; Score 74.8; DB 8; Length 115218;
Best Local Similarity 43.6%; Pred. No. 0.0013;
Matches 428; Conservative 0; Mismatches 552; Indels 2; Gaps 2;



RESULT 7
US-10-473-126-240
; Sequence 240, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 240
; LENGTH: 8056
; TYPE: DNA

```
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-240

  Query Match             6.0%;  Score 71;  DB 8;  Length 8056;
  Best Local Similarity   45.0%;  Pred. No. 0.0026;
  Matches  438;  Conservative    0;  Mismatches  520;  Indels   15;  Gaps    4;

Qy         16 AAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATA 75
              || ||  ||||| | ||  | || |   ||||   | |  | || | |      | | ||
Db         56 AATTTTTTTTTTATTTTTTTTTTAATTTTTAATTTTTATATTTTTTTATATATTTAATT 115

Qy         76 GATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAG 135
                |  |  ||         | | |     |  ||||   || | || || |||| ||| |
Db        116 ATTATTTTTAATTTTTTATATTTTTTTTATTTATTTTTTATTTTTTTATTTATTTTTTAT 175

Qy        136 ATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAA 195
              |||||||   |||||| |||   | |  |||||||  |   | |||| |    ||||
Db        176 TATTTTTTTTAATTATTAATTTTAATTATTATTATAATTTTTAAAAAATATTTTAAAATT 235

Qy        196 TTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATT 255
              ||||           |||| ||  ||    | ||      |   ||    |||
Db        236 TTTAAAAATTTAAAAATTTAAAATTTTAAATTTATTTAATAAAAAAAAATTTTTATT 295

Qy        256 GCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGG 315
              |  | ||   |||||| ||||| ||   ||| || |  | |  |   |
Db        296 TTT--TTTAAAATTATTTTTTAAAAAAATAATTAATAAAAATAAATAAAAATAATAATAA 353

Qy        316 GCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGT 375
              |||  | ||  |    ||  | ||  ||| |   |||  |  |  |  |  |  |
Db        354 TTAAAAAAAAAAAAAAAACGAAATAAAATATTTATAATTTAAA---ATTTAATTTAATAAT 410

Qy        376 AATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACC 435
              |  |  ||  | |   |  || ||   || || |  |||||  |     | |  |
Db        411 ATATTTATTAATTATTTTATTTTTTTTAAAAATAATAAAAAAAAATTAATTTTAAAAAAT 470

Qy        436 ATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTG 495
              |  |||| | ||  ||   | ||       |||||    ||     | | || |  | ||
Db        471 AAAATATATTTAAATAAAAAAAATTAAAATTTATAATATTTATAATATTAAATATTATAT 530

Qy        496 CTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAA 555
              | |   || ||  | |||  | |     ||| |  | |    | |    | ||  | ||
Db        531 ATTAAAATTAAAATTTTATATAATTATTATAAACGTTTTATTAAAATTATTTATATTTAA 590

Qy        556 ATAAAA-----AGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTT 610
              |  ||     | || || |   ||| |    ||  |   ||  |   |  || ||
Db        591 TAATAATATTTATAAAAAAATATTATATAAATTTTAAAATTTTTTAAAAATAATTTTTAA 650

Qy        611 ATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAA 670
              || ||| ||||  | || |   | || || |  |   ||| || ||| | || |   | |
Db        651 ATATAAATTTATTAAAAAAAAAAAAAAAAATAATAAATATAAAAAAATTTTTTTATATTTA 710

Qy        671 TTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCTATTA 730
              | |   || | ||||    |  ||  | |     | || || |   |  |     |||
Db        711 AAATTTAATTTAAAATAATTTATATTATAATT-----AAATTTTTTAATTTTAAAATTTA 765

Qy        731 AGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATAATAAATAGAA 790
              |  |  |||         | |    | |     || ||||   |  ||  | ||
Db        766 ATTAAAATTATAAAATTATATTTAATTATATATTTAAAATAAATTTATAAACGATTATTT 825

Qy        791 TACAGGGAATAAAGCATCAAGACTTATTTTATGGAGAACGGTTACATAAACAAGTATTTG 850
              | |    | |  |  || |  |||| |     ||||  || | || | |
Db        826 AAAATATTTTTATTAAAAAAAATCGTTTTTTTTTATTTTTTTTTACGTATAAAAATTTCGT 885

Qy        851 GAGACATGGGAGCAAATTTTTTTATCAGTTACTACGTATGGAGCAGAACTGTTAGTTTTTT 910
              |    |   |  | |||| || | || |  ||| |    | || | ||  ||
Db        886 AATTATTTAAATTTATATTTTAATTATTTTCGTCGTTTTCGTAATAAATTTTTTTTAAAA 945

Qy        911 TTGGTTTTCTCTGTGTATTCATTATCCCTTTAGGGATATATATACCTTTTTATCTTTTAA 970
              |  ||| | | | |  | | |   | |  ||||||     |||| |      |
Db        946 ATAAATTTTTATTTTTTTATTTAAACGAATAAAATATATATTATAATTTTTTTAAAAATA 1005

Qy        971 AGAGAATGAAAAA 983
              |     | |||||
Db       1006 ATTATTTTAAAAA 1018


RESULT 8
US-10-473-126-386
; Sequence 386, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Methods and nucleic acids for the analysis of  hematopoietic cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/473,126
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 386
;   LENGTH: 8056
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-386

  Query Match             5.9%;  Score 70;  DB 8;  Length 8056;
  Best Local Similarity   44.5%;  Pred. No. 0.0041;
  Matches  502;  Conservative    0;  Mismatches  610;  Indels   16;  Gaps    5;

Qy         16 AAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATA 75
              || ||  ||||| | ||  | || |   ||||   | |  | || | |      | | ||
Db         56 AATTTTTTTTTTATTTTTTTTTTAATTTTTAATTTTTATATTTTTTTATATATTTAATT 115

Qy         76 GATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAG 135
                |  |  ||         | | |     |  ||||   || | || || |||| ||| |
Db        116 ATTATTTTTAATTTTTTATATTTTTTTTATTTATTTTTTATTTTTTTATTTATTTTTTAT 175

Qy        136 ATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAA 195
              |||||||   |||||| |||   | |  |||||||  |   | |||| |    ||||
Db        176 TATTTTTTTTAATTATTAATTTTAATTATTATTATAATTTTTAAAAAATATTTTAAAATT 235

Qy        196 TTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATT 255
              ||||           |||| ||  ||    | ||      |   ||    |||
Db        236 TTTAAAAATTTAAAAATTTTAAAATTTTTAAATTTATTTTAATAAAAAAAAATTTTTATT 295

Qy        256 GCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGG 315
              |  | ||   |||||| ||||| ||   ||| || |  | |  |   |
Db        296 TTT--TTTAAAATTATTTTTTAAAAAAATAATTAATAAAAATAAATAAAAATAATAATAA 353

Qy        316 GCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGT 375
              |||  | ||  |    ||  | ||  ||| |   |||  |  |  |  |  |  |
Db        354 TTAAAAAAAAAAAAAAAACGAAATAAAATATTTATAATTTAAA---ATTTAATTTAATAAT 410

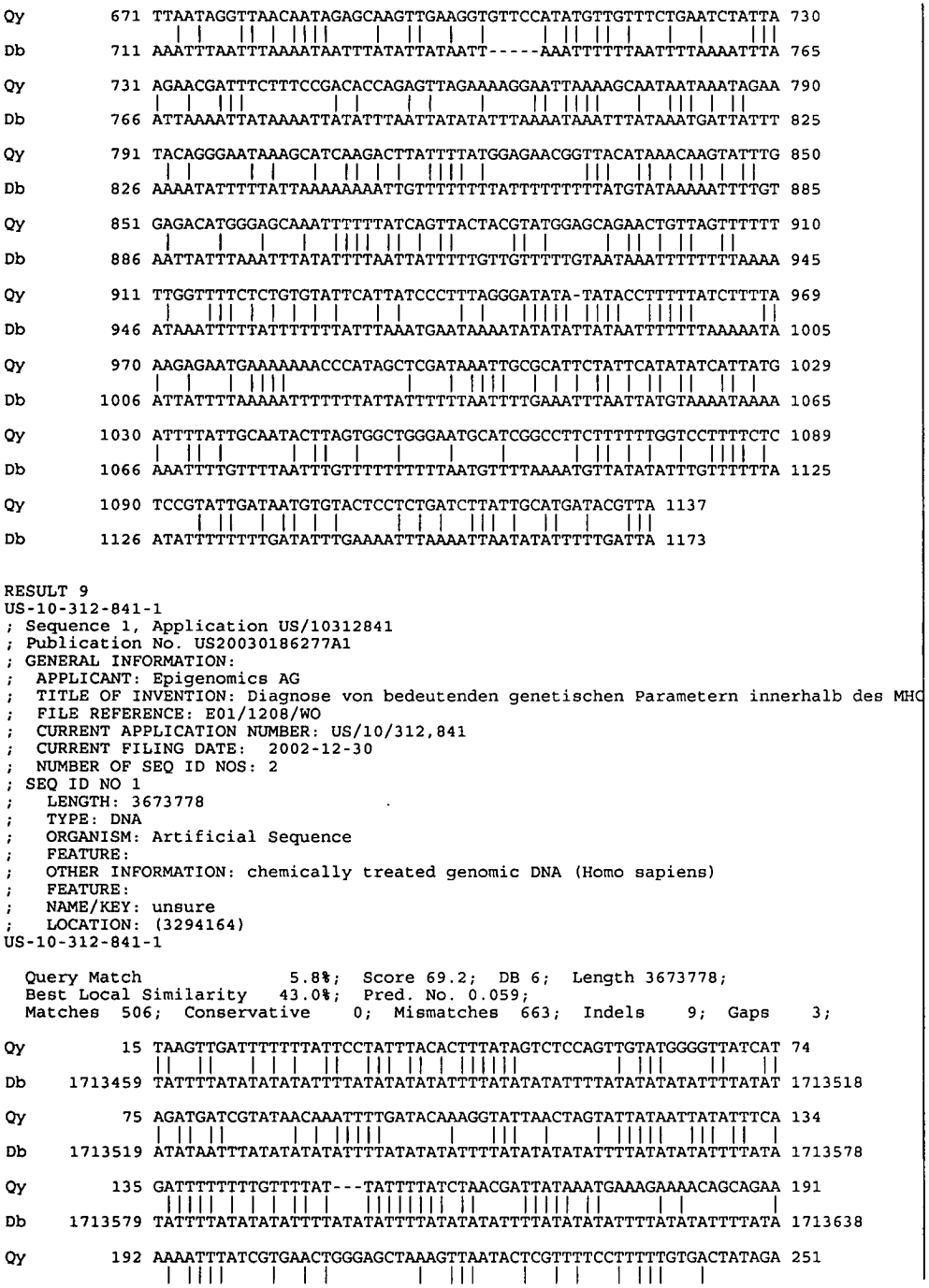
Qy        376 AATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACC 435
              |  |  ||  | |   |  || ||   || || |  |||||  |     | |  |
Db        411 ATATTTATTAATTATTTTATTTTTTTTAAAAATAATAAAAAAAAATTAATTTTAAAAAAT 470

Qy        436 ATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTG 495
              |  |||| | ||  ||   | ||       |||||    ||     | | || |  | ||
Db        471 AAAATATATTTAAATAAAAAAAATTAAAATTTATAATATTTATAATATTAAATATTATAT 530

Qy        496 CTCATTCTTCAATATATATTGATTACCAATA-----TCCTGTCAAAGGATAACCGTAATC 550
              | |   || ||  | |||  | |     |||       |     | |||  ||      |
Db        531 ATTAAAATTAAAATTTTATATAATTATTATAAATGTTTTATTAAAATTATTTATATTTAA 590

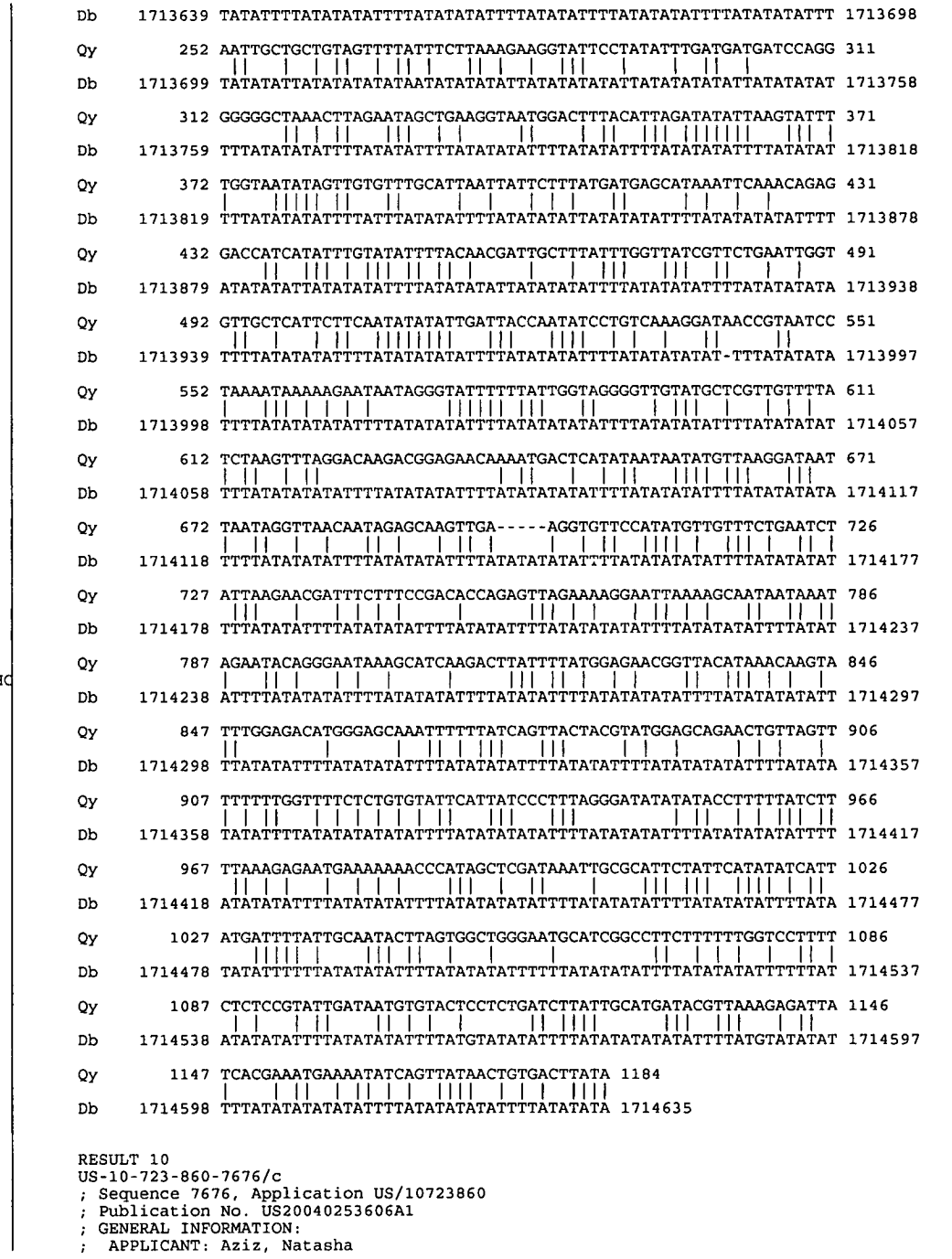
Qy        551 CTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTT 610
              | ||||    | || || |   ||| |    ||  |   ||  |   |  || ||
Db        591 TAATAATATTTATAAAAAAATATTATATAAATTTTAAAATTTTTTAAAAATAATTTTTAA 650

Qy        611 ATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAA 670
              || ||| ||||  | || |   | || || |  |   ||| || ||| | || |   | |
Db        651 ATATAAATTTATTAAAAAAAAAAAAAAAAATAATAAATATAAAAAAATTTTTTTATATTTA 710
```



RESULT 9
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 5.8%; Score 69.2; DB 6; Length 3673778;
Best Local Similarity 43.0%; Pred. No. 0.059;
Matches 506; Conservative 0; Mismatches 663; Indels 9; Gaps 3;



RESULT 10
US-10-723-860-7676/c
; Sequence 7676, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7676
; LENGTH: 6816
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1919)..(1989)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3223)..(3247)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6189)..(6189)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6209)..(6209)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6215)..(6221)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6223)..(6230)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6234)..(6239)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6241)..(6244)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6248)..(6248)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6250)..(6250)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6254)..(6258)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6261)..(6261)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6276)..(6276)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6278)..(6281)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6291)..(6300)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6305)..(6305)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6307)..(6307)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6322)..(6322)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6324)..(6327)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6338)..(6339)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6341)..(6342)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6351)..(6351)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6386)..(6386)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6388)..(6388)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6413)..(6414)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6416)..(6416)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6418)..(6418)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6434)..(6435)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6438)..(6438)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6463)..(6463)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6465)..(6465)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6470)..(6470)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6483)..(6483)

```
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6485)..(6485)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6494)..(6494)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6517)..(6517)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6534)..(6536)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6557)..(6557)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6562)..(6563)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6607)..(6607)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6619)..(6620)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6642)..(6642)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6644)..(6646)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6660)..(6660)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6663)..(6663)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:

  Query Match             5.8%;  Score 68.6;  DB 8;  Length 6816;
  Best Local Similarity   39.2%;  Pred. No. 0.0071;
  Matches  212;  Conservative    0;  Mismatches  324;  Indels    5;  Gaps    1;

Qy         85 ATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTTTTTTT 144
Db       6707 AAAAAATATTNNANAAAANNANTTTTNAANATTNNNNNNNNNATNTTNTATTTTTTTATA 6648

Qy        145 GTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTATCGTG 204
Db       6647 TNNNANTAAAAAAAAAAAAAATAATATANNAAAAAAAAAAAANAAAAAAAAAAAATTTTTTT 6588

Qy        205 AACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGTA 264
Db       6587 TATAAATTTTTTTTAAAAAAAAAANNTTTTNTTTTAAAAAATATTTAAATANNNTTTTTA 6528

Qy        265 GTTTTATTTCTTAAAGAAGGTATTCCTATATTTG-----ATGATGATCCAGGGGGGGCTA 319
Db       6527 TTTTTAATTTNTTTATTTTAAAAATAAAAATATNTAAAAAAANTNATTTTTTTTAAANAA 6468

Qy        320 AACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATA 379
```

```
Db       6467 AANTNAAAATATTTTTTTTTTATTTTTTTNCCNNAAAAAAATAATAATAANTNANNTAAAA 6408

Qy        380 TAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCA 439
Db       6407 AATTTTTTTTTTAAATAAANTNTTTTTATAAAAAATATTTTTTTTAAAAAAAAAAANTAA 6348

Qy        440 TATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCA 499
Db       6347 AATTTNNTNNTTTTTTTATTNNNNTNTTATTTATTTTTTTNTNTTTTNNNNNNNNNTTT 6288

Qy        500 TTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATAA 559
Db       6287 TTTTTTNNNNTNTTTTTTTTTTTTTTTNTTNNNNNTTTNTNTTTNNNNTNNNNNNTTTNNN 6228

Qy        560 AAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGTT 619
Db       6227 NNNNNTNNNNNNNTTTTTNTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTT 6168

Qy        620 T 620
Db       6167 T 6167
```

```
RESULT 11
US-10-311-455-74
; Sequence 74, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 74
;   LENGTH: 5728
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-74

  Query Match             5.8%;  Score 68.4;  DB 6;  Length 5728;
  Best Local Similarity   48.8%;  Pred. No. 0.0072;
  Matches  244;  Conservative    0;  Mismatches  251;  Indels    5;  Gaps    2;

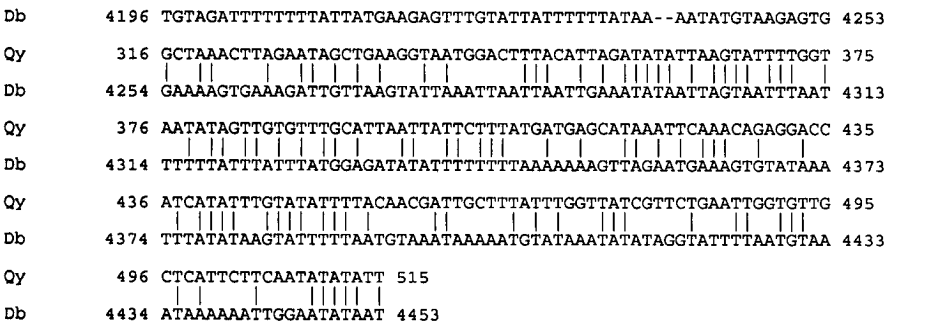
Qy         19 TTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATAGAT 78
Db       3956 TTTATGTTATTTTTTATTTTGATAGGTTATATTTTTGATTTTTTTTTTTTTAAAGATAAA 4015

Qy         79 GATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTA---TTATAATTATATTTCAG 135
Db       4016 GGGAAATATATTTATGGAAATGTAATAATATTTAATAATATTTTTAATATTTTATAAGAG 4075

Qy        136 ATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAA 195
Db       4076 ATTTTTTATAAATAATTATTTTATTTAGAGTTATTTTAAAATATAAGAAATTTGGATTTG 4135

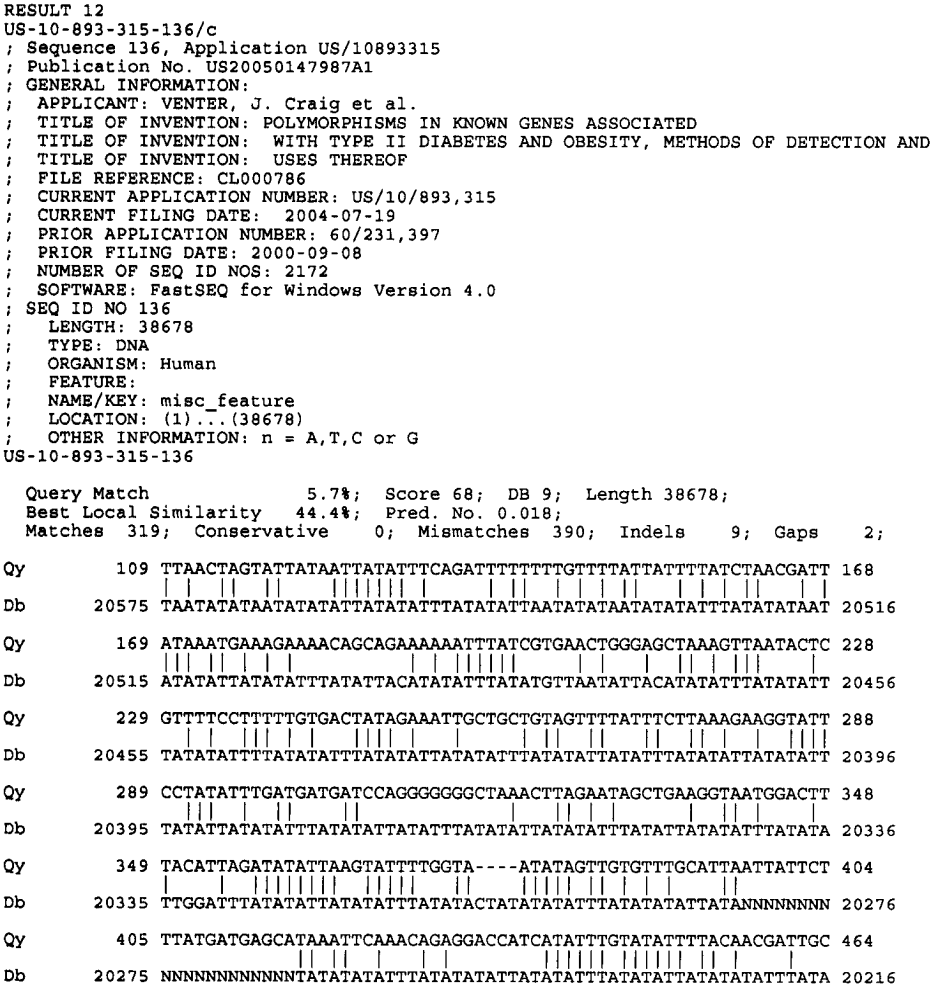
Qy        196 TTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATT 255
Db       4136 TTTGGAGAGAAAATTGGGGATAGGAAAATTTTATTATTTTTTTATTATAAAATAAATATA 4195

Qy        256 GCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGG 315
```



RESULT 12
US-10-893-315-136/c
; Sequence 136, Application US/10893315
; Publication No. US20050147987A1
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH TYPE II DIABETES AND OBESITY, METHODS OF DETECTION AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000786
; CURRENT APPLICATION NUMBER: US/10/893,315
; CURRENT FILING DATE: 2004-07-19
; PRIOR APPLICATION NUMBER: 60/231,397
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 2172
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 136
; LENGTH: 38678
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(38678)
; OTHER INFORMATION: n = A,T,C or G
US-10-893-315-136

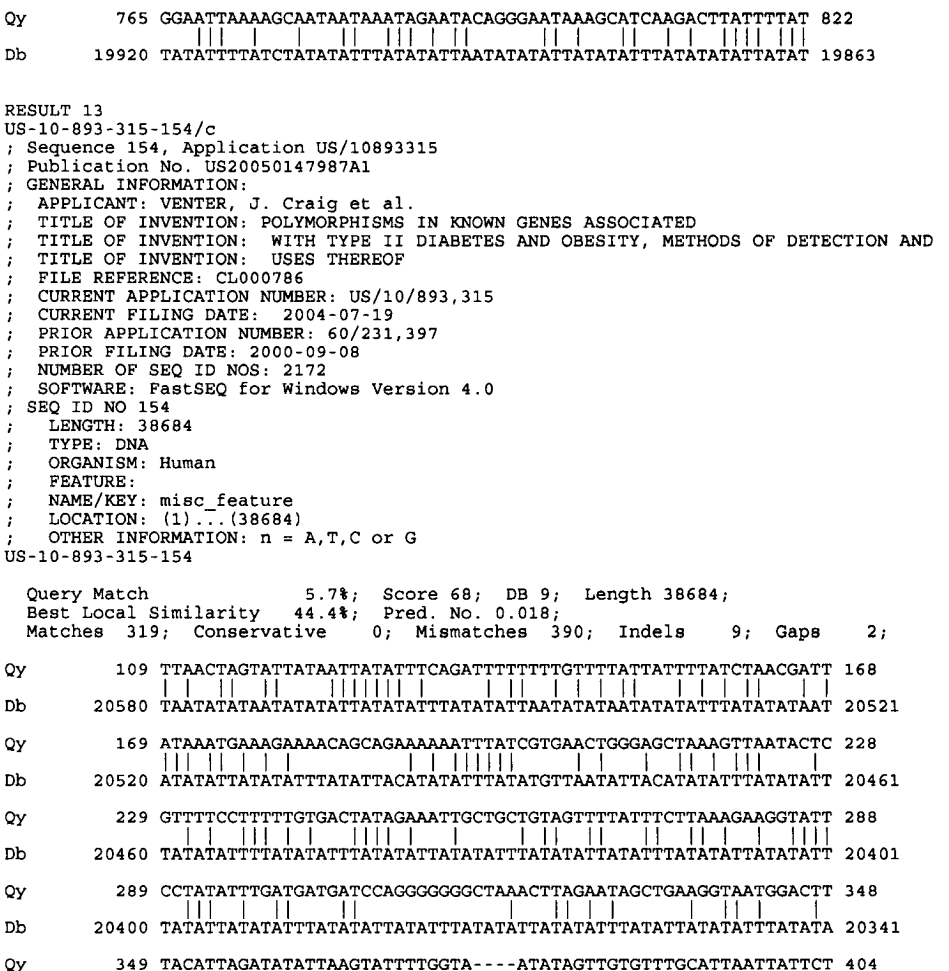
Query Match 5.7%; Score 68; DB 9; Length 38678;
Best Local Similarity 44.4%; Pred. No. 0.018;
Matches 319; Conservative 0; Mismatches 390; Indels 9; Gaps 2;

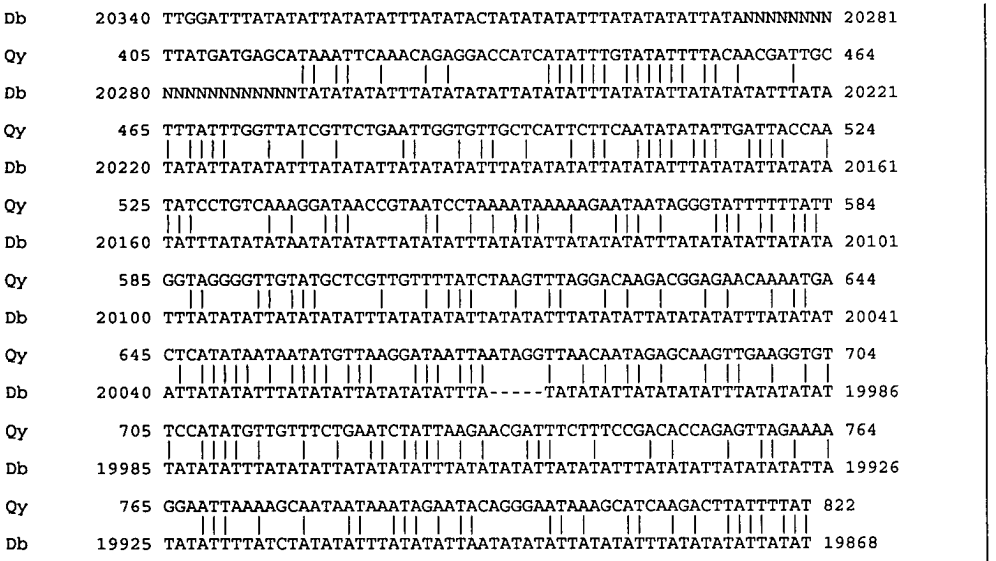




RESULT 13
US-10-893-315-154/c
; Sequence 154, Application US/10893315
; Publication No. US20050147987A1
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH TYPE II DIABETES AND OBESITY, METHODS OF DETECTION AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000786
; CURRENT APPLICATION NUMBER: US/10/893,315
; CURRENT FILING DATE: 2004-07-19
; PRIOR APPLICATION NUMBER: 60/231,397
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 2172
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 154
; LENGTH: 38684
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(38684)
; OTHER INFORMATION: n = A,T,C or G
US-10-893-315-154

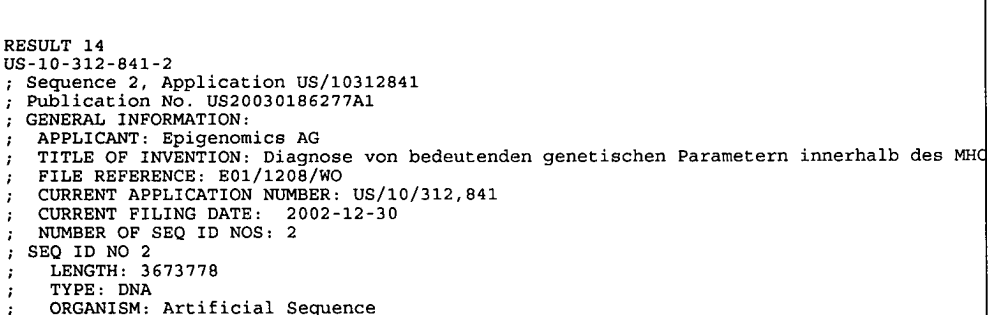
Query Match 5.7%; Score 68; DB 9; Length 38684;
Best Local Similarity 44.4%; Pred. No. 0.018;
Matches 319; Conservative 0; Mismatches 390; Indels 9; Gaps 2;

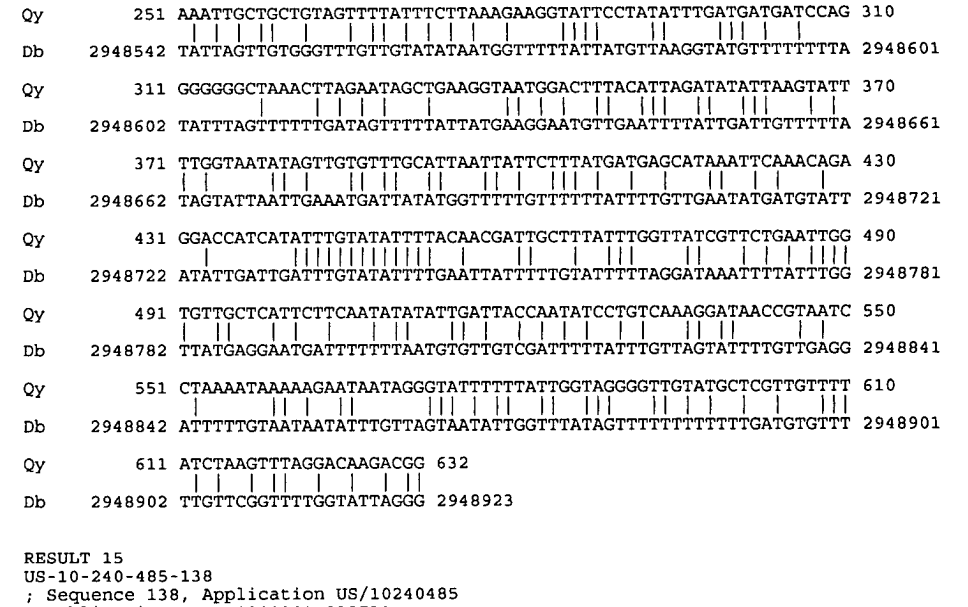




RESULT 14
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

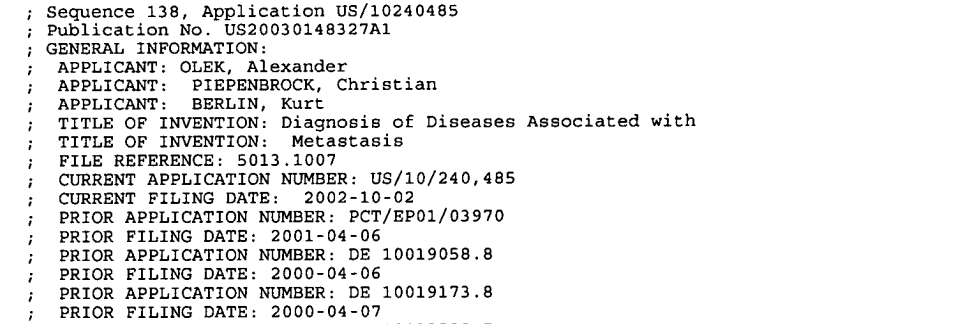
Query Match 5.7%; Score 68; DB 6; Length 3673778;
Best Local Similarity 45.5%; Pred. No. 0.1;
Matches 283; Conservative 0; Mismatches 335; Indels 4; Gaps 1;

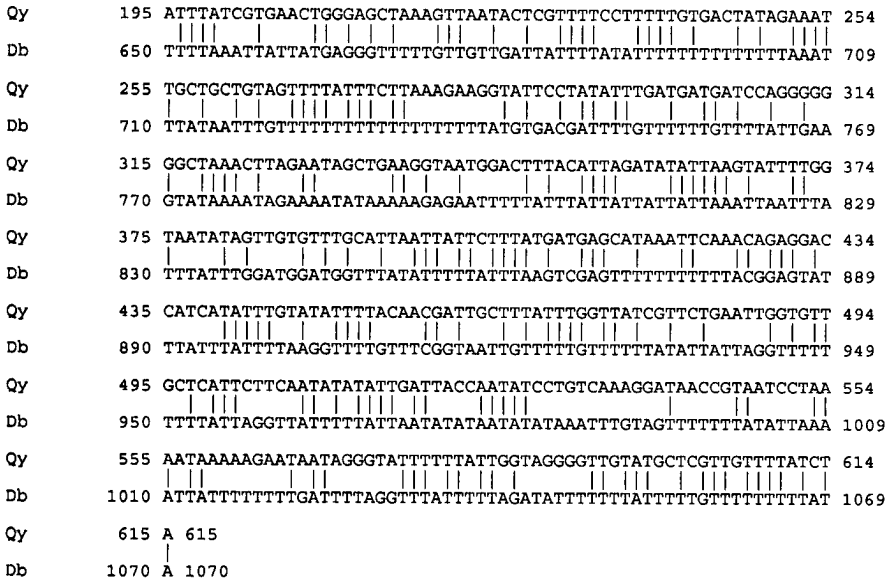




RESULT 15
US-10-240-485-138
; Sequence 138, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: Metastasis
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240,485
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 138
; LENGTH: 14551
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-138

Query Match 5.7%; Score 67.8; DB 6; Length 14551;
Best Local Similarity 46.6%; Pred. No. 0.013;
Matches 252; Conservative 0; Mismatches 287; Indels 2; Gaps 1;





Search completed: May 11, 2006, 07:09:05
Job time : 793.135 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 05:23:58 ; Search time 612.35 Seconds
(without alignments)
7892.621 Million cell updates/sec

Title: US-09-423-093B-2_COPY_858_2042
Perfect score: 1185
Sequence: 1 gtgaagtcagcggctaagtt.....gttataactgtgacttataa 1185

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/1/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/1/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	64.8	5.5	28536	17	US-11-011-332A-152	Sequence 152, App
	2	63.2	5.3	28536	17	US-11-011-332A-156	Sequence 156, App
c	3	61.2	5.2	173602	17	US-11-121-086-25	Sequence 25, Appl
	4	60.4	5.1	173602	17	US-11-121-086-25	Sequence 25, Appl
	5	58.8	5.0	6801	10	US-10-240-708-62	Sequence 62, Appl
	6	57	4.8	171486	17	US-11-121-086-105	Sequence 105, App
c	7	56.4	4.8	171486	17	US-11-121-086-105	Sequence 105, App
c	8	55.2	4.7	1457619	17	US-11-098-686-8739	Sequence 8739, Ap
c	9	54.8	4.6	26772	10	US-10-995-561-13313	Sequence 13313, A
c	10	54.8	4.6	54946	10	US-10-995-561-13479	Sequence 13479, A
	11	54.6	4.6	9001	17	US-11-011-332A-2	Sequence 2, Appli
	12	54.6	4.6	28536	17	US-11-011-332A-151	Sequence 151, App
	13	53.2	4.5	596	7	US-09-925-065A-436829	Sequence 436829,
	14	53.2	4.5	597	12	US-10-301-480-498452	Sequence 498452,
	15	53.2	4.5	597	12	US-10-301-480-1111861	Sequence 1111861,
	16	53.2	4.5	6070	10	US-10-240-708-10	Sequence 10, Appl
	17	53	4.5	9001	17	US-11-011-332A-4	Sequence 4, Appli
	18	53	4.5	28536	17	US-11-011-332A-155	Sequence 155, App
	19	53	4.5	156735	17	US-11-121-086-93	Sequence 93, Appl
c	20	52.8	4.5	674	12	US-10-301-480-532538	Sequence 532538,
c	21	52.8	4.5	674	12	US-10-301-480-1145947	Sequence 1145947,
	22	51.6	4.4	6456	17	US-11-011-332A-92	Sequence 92, Appl
	23	51.4	4.3	10467	10	US-10-240-708-2	Sequence 2, Appli
c	24	51.2	4.3	1791	10	US-10-750-185-43115	Sequence 43115, A
c	25	51.2	4.3	1791	10	US-10-750-623-43115	Sequence 43115, A
	26	51	4.3	587	12	US-10-301-480-572793	Sequence 572793,
	27	51	4.3	587	12	US-10-301-480-1186202	Sequence 1186202,
	28	51	4.3	648	12	US-10-301-480-572792	Sequence 572792,
	29	51	4.3	648	12	US-10-301-480-1186201	Sequence 1186201,
	30	51	4.3	8093	10	US-10-240-708-31	Sequence 31, Appl
	31	50.4	4.3	8961	10	US-10-240-708-80	Sequence 80, Appl
	32	50.2	4.2	984	12	US-10-301-480-607658	Sequence 607658,
	33	50.2	4.2	984	12	US-10-301-480-607659	Sequence 607659,
	34	50.2	4.2	984	12	US-10-301-480-1221067	Sequence 1221067,
	35	50.2	4.2	984	12	US-10-301-480-1221068	Sequence 1221068,
	36	50.2	4.2	2480	7	US-09-925-065A-687693	Sequence 687693,
	37	50.2	4.2	2480	7	US-09-925-065A-687694	Sequence 687694,
	38	50.2	4.2	2480	7	US-09-925-065A-687695	Sequence 687695,
	39	50.2	4.2	11131	10	US-10-240-708-27	Sequence 27, Appl
	40	50.2	4.2	49979	10	US-10-995-561-13443	Sequence 13443, A
	41	49.8	4.2	6866	10	US-10-240-708-20	Sequence 20, Appl
	42	49.8	4.2	26772	10	US-10-995-561-13313	Sequence 13313, A
	43	49.8	4.2	54946	10	US-10-995-561-13479	Sequence 13479, A
	44	49.2	4.2	5610	10	US-10-240-708-54	Sequence 54, Appl
c	45	49	4.1	991	12	US-10-301-480-558212	Sequence 558212,

ALIGNMENTS

RESULT 1
US-11-011-332A-152
; Sequence 152, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
; APPLICANT: Foekens, John
; APPLICANT: Harbeck, Nadia
; APPLICANT: Koenig, Thomas
; APPLICANT: Maier, Sabine
; APPLICANT: Martens, John
; APPLICANT: Model, Fabian
; APPLICANT: Nimmrich, Inko
; APPLICANT: Rujan, Tamas
; APPLICANT: Schmitt, Manfred
; APPLICANT: Lesche, Ralf
; APPLICANT: Dietrich, Dimo
; APPLICANT: Mueller, Volkmar
; APPLICANT: Kluth, Antje
; APPLICANT: Schwope, Ina
; APPLICANT: Hartmann, Oliver
; APPLICANT: Adorjan, Peter
; TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SU
; TITLE OF INVENTION: BREAST CELL PROLIFERATIVE DISORDER PATIENTS
; FILE REFERENCE: 47675-99
; CURRENT APPLICATION NUMBER: US/11/011,332A
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 10/517,741
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: PCT/EP2003/010881
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: DE 10245779.4
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: DE 10300096.8
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: DE 10317955.0
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: PCT/EP2004/014170

; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: EP 03090432.0
; PRIOR FILING DATE: 2003-12-11
; PRIOR APPLICATION NUMBER: EP 04090041.7
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: EP 04090380.9
; PRIOR FILING DATE: 2004-09-30
; PRIOR APPLICATION NUMBER: EP 04090127.4
; PRIOR FILING DATE: 2004-04-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 152
; LENGTH: 28536
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-152

Query Match 5.5%; Score 64.8; DB 17; Length 28536;
Best Local Similarity 46.7%; Pred. No. 0.24;
Matches 240; Conservative 0; Mismatches 272; Indels 2; Gaps 1;

```
Qy        109 TTAACTAGTATTATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATT 168
              |||| ||||  | | |   |   | ||  ||| |   |||||  |   | |  || ||||
Db       1016 TTAAATAGTGATGTTAGATTTGATAAGGATTTGTAAATTTTAAAAGAGTGTAAAAAGATT 1075

Qy        169 ATAAATGAAAGAAAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTC 228
              ||| | |||  | | |           ||||||  |  |           |||      || |
Db       1076 ATAGAAGAATAATATTATATTTTTGTATTTATAGAAATGGTTAGTTTAAGAGATGTATTT 1135

Qy        229 GTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATT 288
                |||    |   ||    |   | ||  ||   | |||  |      | ||  |      ||
Db       1136 AGTTTAGGGTGGTTGTAAGTTTTATTTTTTGAATTTGTTATTAGAAGTTAAAAGAAATTT 1195

Qy        289 CCTATATTTGATGATGATCCAGGGGGGGCTAAACT--TAGAATAGCTGAAGGTAATGGAC 346
              ||||  ||| |  ||||  |          |  |  ||  |   |  |||  | |
Db       1196 TGTATAATTGTTTTTGATGGAAATATAAATTGGTTGATGTAAATGAAATATATAAAGTAG 1255

Qy        347 TTTACATTAGATATATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTT 406
              ||   ||  || ||||      ||   |||  | ||    |     |  ||||||   ||
Db       1256 TTGGTGTTTTATTTATTTTTATAATTATAAATGAAATTAAATGATTAAAAATTATAGATT 1315

Qy        407 ATGATGAGCATAAATTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTT 466
              ||  ||   |  || |  ||  |  ||   |||||  | |||||  |  | |  ||
Db       1316 TTGGGGATTTTTTTTTTATTGAGGAGTTTATGGAATTTGTTTTTTTTAGTATTAATAATT 1375

Qy        467 TATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATA 526
              ||| ||    || |    |    ||    | || ||    || |||| ||||
Db       1376 GTGTTGATTTATTTTTTTTTGTTTAATTTTGTATATATTAAAATTAGGTGGTTACGAATA 1435

Qy        527 TCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGG 586
              |  |  ||  ||||    ||  ||| | ||  |   ||| |   | || ||| ||
Db       1436 AAATTTAGAAATATAATTTATATTTTAATAAAATGATTTTAAAATTATTTTATTTTTTAT 1495

Qy        587 TAGGGGTTGTATGCTCGTTGTTTTATCTAAGTTT 620
              |  || ||   | | ||| | | || || ||||
Db       1496 TGTGGTTTTATTTGTTGTTTTATAATGTAGGTTT 1529
```

RESULT 2
US-11-011-332A-156
; Sequence 156, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
; APPLICANT: Foekens, John
; APPLICANT: Harbeck, Nadia
; APPLICANT: Koenig, Thomas
; APPLICANT: Maier, Sabine
; APPLICANT: Martens, John
; APPLICANT: Model, Fabian
; APPLICANT: Nimmrich, Inko
; APPLICANT: Rujan, Tamas
; APPLICANT: Schmitt, Manfred
; APPLICANT: Lesche, Ralf
; APPLICANT: Dietrich, Dimo
; APPLICANT: Mueller, Volkmar
; APPLICANT: Kluth, Antje
; APPLICANT: Schwope, Ina
; APPLICANT: Hartmann, Oliver
; APPLICANT: Adorjan, Peter
; TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SUR
; TITLE OF INVENTION: BREAST CELL PROLIFERATIVE DISORDER PATIENTS
; FILE REFERENCE: 47675-99
; CURRENT APPLICATION NUMBER: US/11/011,332A
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 10/517,741
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: PCT/EP2003/010881
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: DE 10245779.4
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: DE 10300096.8
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: DE 10317955.0
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: PCT/EP2004/014170
; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: EP 03090432.0
; PRIOR FILING DATE: 2003-12-11
; PRIOR APPLICATION NUMBER: EP 04090041.7
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: EP 04090380.9
; PRIOR FILING DATE: 2004-09-30
; PRIOR APPLICATION NUMBER: EP 04090127.4
; PRIOR FILING DATE: 2004-04-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 156
; LENGTH: 28536
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-156

Query Match 5.3%; Score 63.2; DB 17; Length 28536;
Best Local Similarity 46.5%; Pred. No. 0.44;
Matches 239; Conservative 0; Mismatches 273; Indels 2; Gaps 1;

```
Qy        109 TTAACTAGTATTATAATTATATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATT 168
              |||| ||||  | | |   |   | ||  ||| |   |||||  |   | |  || ||||
Db       1016 TTAAATAGTGATGTTAGATTTGATAAGGATTTGTAAATTTTAAAAGAGTGTAAAAAGATT 1075

Qy        169 ATAAATGAAAGAAAACAGCAGAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTC 228
              ||| | |||  | | |           ||||||  |  |           |||      || |
Db       1076 ATAGAAGAATAATATTATATTTTTGTATTTATAGAAATGGTTAGTTTAAGAGATGTATTT 1135

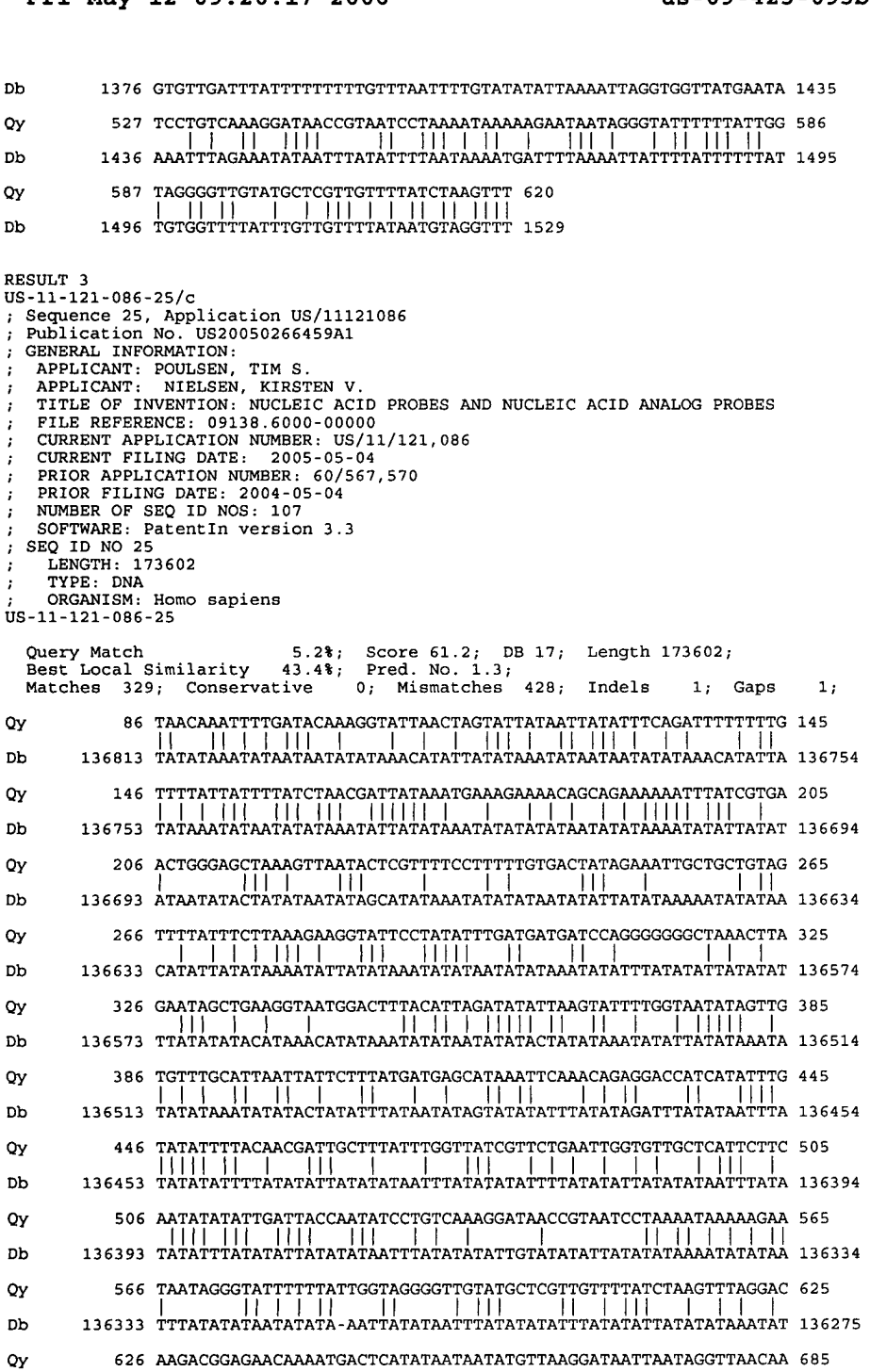
Qy        229 GTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATT 288
                |||    |   ||    |   | ||  ||   | |||  |      | ||  |      ||
Db       1136 AGTTTAGGGTGGTTGTAAGTTTTATTTTTTGAATTTGTTATTAGAAGTTAAAAGAAATTT 1195

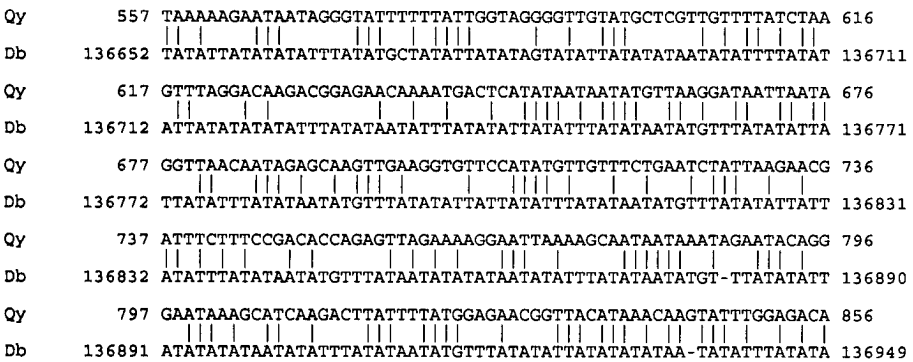
Qy        289 CCTATATTTGATGATGATCCAGGGGGGGCTAAACT--TAGAATAGCTGAAGGTAATGGAC 346
              ||||  ||| |  ||||  |          |  |  ||  |   |  |||  | |
Db       1196 TGTATAATTGTTTTTGATGGAAATATAAATTGGTTGATGTAAATGAAATATATAAAGTAG 1255

Qy        347 TTTACATTAGATATATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTCTTT 406
              ||   ||  || ||||      ||   |||  | ||    |     |  ||||||   ||
Db       1256 TTGGTGTTTTATTTATTTTTATAATTATAAATGAAATTAAATGATTAAAAATTATAGATT 1315

Qy        407 ATGATGAGCATAAATTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTT 466
              ||  ||   |  || |  ||  |  ||   |||||  | |||||  |  | |  ||
Db       1316 TTGGGGATTTTTTTTTTATTGAGGAGTTTATGGAATTTGTTTTTTTTAGTATTAATAATT 1375

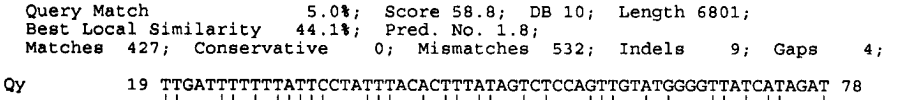
Qy        467 TATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATA 526
              ||| ||    || |    |    ||    | || ||    || |||| ||||
```


Db 1376 GTGTTGATTTATTTTTTTTTGTTTAATTTTGTATATATTAAAATTAGGTGGTTATGAATA 1435
Qy 527 TCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGG 586
Db 1436 AAATTTAGAAATATAATTTATATTTTAATAAAATGATTTTAAAATTATTTTATTTTTTAT 1495
Qy 587 TAGGGGTTGTATGCTCGTTGTTTTATCTAAGTTT 620
Db 1496 TGTGGTTTTATTTGTTGTTTTATAATGTAGGTTT 1529
RESULT 3
US-11-121-086-25/c
; Sequence 25, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; LENGTH: 173602
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-25
Query Match 5.2%; Score 61.2; DB 17; Length 173602;
Best Local Similarity 43.4%; Pred. No. 1.3;
Matches 329; Conservative 0; Mismatches 428; Indels 1; Gaps 1;
Qy 86 TAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTTTTTTTG 145
Db 136813 TATATAAATATAATAATATATAAACATATTATATAAATATAATAATATATAAACATATTA 136754
Qy 146 TTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTATCGTGA 205
Db 136753 TATAAATATAATATATAAATATTATATAAATATATATATAATATATAAAATATATTATAT 136694
Qy 206 ACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGTAG 265
Db 136693 ATAATATACTATATAATATAGCATATAAATATATATAATATATTATATAAAAATATATAA 136634
Qy 266 TTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTAAACTTA 325
Db 136633 CATATTATATAAAATATTATATAAATATATAATATATAAATATATTTATATATTATATAT 136574
Qy 326 GAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATATAGTTG 385
Db 136573 TTATATATACATAAACATATAAATATATAATATATACTATATAAATATATTATATAAATA 136514
Qy 386 TGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCATATTTG 445
Db 136513 TATATAAATATATACTATATTTATAATATAGTATATATTTATATAGATTTATATAATTTA 136454
Qy 446 TATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCTTC 505
Db 136453 TATATATTTTATATATTATATATAATTTATATATTTTATATATTATATATAATTTATA 136394
Qy 506 AATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAA 565
Db 136393 TATATTTATATATTATATATAATTTATATATATTGTATATATTATATATAAAATATATAA 136334
Qy 566 TAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGTTTAGGAC 625
Db 136333 TTTATATATAATATATA-AATTATATAATTTATATATATTTATATATTATATATAAATAT 136275
Qy 626 AAGACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAATTAATAGGTTAACAA 685
Db 136274 ATATAATTTATATAATTTATATATTATATATATAAATATATAATTTATATATATTTATAT 136215
Qy 686 TAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCTATTAAGAACGATTTCTTTC 745
Db 136214 AATATATATTATATATTTATATATATAATATATATTTATATTTATATATCTATATAATAT 136155
Qy 746 CGACACCAGAGTTAGAAAAGGAATTAAAAGCAATAATAAATAGAATACAGGGAATAAAGC 805
Db 136154 ATATTATATATTTATATATTTATAAATATATAATATATATTTATATATAAAAATATATAC 136095
Qy 806 ATCAAGACTTATTTTATGGAGAACGGTTACATAAACAA 843
Db 136094 ATATATAATTTATATATATATTACATATACATATAAAA 136057
RESULT 4
US-11-121-086-25
; Sequence 25, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; LENGTH: 173602
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-25
Query Match 5.1%; Score 60.4; DB 17; Length 173602;
Best Local Similarity 45.8%; Pred. No. 1.8;
Matches 435; Conservative 0; Mismatches 501; Indels 14; Gaps 6;
Qy 87 AACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTTTTTTTGT 146
Db 136173 AATATAAATATATATTATATATATAAATATATAATATATTATATAAATATATATAAAT 136232
Qy 147 TTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTATCGTGAA 206
Db 136233 TATATATTTATATATATAATATATAAATTATATAAATTATATATATTTATATATAATATA 136292
Qy 207 CTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGTAGT 266
Db 136293 TAAATATATATAAATTATATAATTTATATATTATATATAAATTATATATTTTATATATAA 136352
Qy 267 TTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGGGGCTAAACTTAG 326
Db 136353 TATATACAATATATATAAATTATATATAATATATAAATATATATAAATTATATATAATAT 136412
Qy 327 AATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAATATAGTTGT 386
Db 136413 ATAAAATATATATAAATTA-TATATAATATATAAAATATATATAAATTATATAAATCTAT 136471
Qy 387 GTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCATATTTGT 446
Db 136472 ATAAATATATACTATATTATAAATATAGTATATATTTATATATATTTATATAATATATTT 136531
Qy 447 ATATTTTACA-ACGATTGCTTTATTTGGTTATCGTTCT---------GAATTGGTGTTGC 496
Db 136532 ATATAGTATATATTATATATTTATATGTTTATGTATATATAAATATATAATATATAAATA 136591
Qy 497 TCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAA 556
Db 136592 TATTTATATATTATATATTTATATAATATTTTATATAATATGTTATATATTTTTATATAA 136651



RESULT 5
US-10-240-708-62
; Sequence 62, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/240,708
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 62
; LENGTH: 6801
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-62

Query Match 5.0%; Score 58.8; DB 10; Length 6801;
Best Local Similarity 44.1%; Pred. No. 1.8;
Matches 427; Conservative 0; Mismatches 532; Indels 9; Gaps 4;





RESULT 6
US-11-121-086-105
; Sequence 105, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES

```
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 105
;   LENGTH: 171486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-105

  Query Match             4.8%;  Score 57;  DB 17;  Length 171486;
  Best Local Similarity   45.2%;  Pred. No. 6.4;
  Matches  415;  Conservative    0;  Mismatches  490;  Indels   14;  Gaps    5;

Qy         91 AATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGATTTTTTTTGTTTTA 150
Db      16086 AAATATAAATATAATATTATATATTATATATATTGTATAATATATATAATATATATTATA 16145

Qy        151 TTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAATTTATCGTGAACTGG 210
Db      16146 TTATATAATATATATCTTATATATTATATATTATATAAGATATAATATTATATGTTATAT 16205

Qy        211 GAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGTAGTTTTA 270
Db      16206 ATTATATATTATATAATATATTATATTATAATTATTATAATTATGTTATATAATTAATAT 16265

Qy        271 TTTCTTAAAGAAGGTAT-TCCTATATTTGATGATGATCCAGGGGGGGCTAAACTTAGAAT 329
Db      16266 TATATAATATATAATATGTATAATATTAATATAATATATAATATATAACATATAATATTT 16325

Qy        330 AGCTGAAGGTAATGGACTTTACATTAGATA-TATTAAGTATTTTGGTAATATAGTTGTGT 388
Db      16326 ATATTAATATAATATATTATATGATATATTATATAACATAATATTTATAATATTAATA 16385

Qy        389 TTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCATATTTGTAT 448
Db      16386 TAATATATAATATTTGATATTTATAAATTATATAATATATAATATTTATAATATTTATAT 16445

Qy        449 ATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCTTCA-A 507
Db      16446 AATATAATATAAATATTATATAAGATTTATATTATATAATAATTATATTACATATTAATA 16505

Qy        508 TATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAATA 567
Db      16506 TATAATTATATTAATATAATTAAAACAATATAATATATAATACATAATATATAATATATT 16565

Qy        568 ATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGTTTAGGACAA 627
Db      16566 ATGTATTATATATTAT--------ATATTTTATATAATATTATATATTATTATATAATAT 16617

Qy        628 GACGGAGAACAAAATGACTCATATAATAATATGTTAAGGATAATTAATAGGTTAACAATA 687
Db      16618 ATTAATAAATATATTATATATATTTATTATATAGTATAATATATTACATATTAATATAAT 16677

Qy        688 GAGCAAGTTGAAGGTGTTCCATATGTTGTTTCTGAATCTATTAAGAACGATTTCTTTCCG 747
Db      16678 AAATATATTAATTATATAATAAATATATTAATATAATATATTATATAATAAATATATTAA 16737

Qy        748 ACACCAGAGTTAGAAAAGGAATTAAAAGCAATAATAAATAGAATACAGGGAATAAAGCAT 807
Db      16738 TTAATATAATATATTATATATGTCTTAAATATATAACATATAATACATTACATATTATAT 16797

Qy        808 CAAGACTTATTTTATGGAGAACGGTTACATAAACAAGTATTTGGAGACATGGGAGCAAAT 867
Db      16798 ATAATTATATTATATAATTATAGATAATACATATATTATCTATAATATATAATATATTAT 16857

Qy        868 ---TTTTTATCAGTTACTACGTATGGAGCAGAACTGTTAGTTTTTTTTGGTTTTCTCTGT 924
Db      16858 ATATTTATATATAAATACATTTATATATTATATATGATAGATATATATAAATATATTTGT 16917

Qy        925 GTATTCATTATCCCTTTAGGGATATATATACCTTTTTATCTTTTAAAGAGAATGAAAAAA 984
Db      16918 ATATTTATATATGATAGATATATATAAATATATTTGTATATTTATATATGATAGATATAT 16977

Qy        985 ACCCATAGCTCGATAAATT 1003
Db      16978 ATAAATATATTTGTATATT 16996


RESULT 7
US-11-121-086-105/c
; Sequence 105, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 105
;   LENGTH: 171486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-105

  Query Match             4.8%;  Score 56.4;  DB 17;  Length 171486;
  Best Local Similarity   44.1%;  Pred. No. 8.1;
  Matches  335;  Conservative    0;  Mismatches  416;  Indels    9;  Gaps    2;

Qy         69 TATCATAGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTAT 128
Db      16851 TATTATATATTATAGATAATATATGTATTATCTATAATTATATAATATAATTATATATAA 16792

Qy        129 ATTTCAGATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCA 188
Db      16791 TATGTAATGTATTATATGTTATATATTTAAGACATATATAATATATTATATTAATTAATA 16732

Qy        189 GAAAAAATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTAT 248
Db      16731 TATTTATTATATAATATATTATATTAATATATTTATTATATAATTAATATAT-TTATTAT 16673

Qy        249 AGAAATTGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCC 308
Db      16672 ATTAATATGTAATATATTATACTATATAATAAATATATATAATATATTTATTAATATATT 16613

Qy        309 AGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTA 368
Db      16612 ATATAATAATATATAATATTATATAAAATATATAATATATAATACATAATATATTATATA 16553

Qy        369 TTTTG--------GTAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAA 420
Db      16552 TTATGTATTATATATTATATTGTTTTAATTATATTAATATAATTATATATTAATATGTAA 16493

Qy        421 TTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGT 480
Db      16492 TATAATTATTATATAATATAAATCTTATATAATATTTATATTATATTATATAAATATTAT 16433

Qy        481 TCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGAT 540
Db      16432 AAATATTATATATTATATAATTTATAAATATCAAATATTATATATTATATTAATATTATA 16373

Qy        541 AACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGC 600
Db      16372 AATATTATGTTATATAATATATCATATAATATATATTATATTAATATAAATATTATATGT 16313

Qy        601 TCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATGACTCATATAATAATATG 660
Db      16312 TATATATTATATATTATATTAATATTATACATATTATATATTATATAATATTAATTATAT 16253
```

```
Qy        661 TTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTGTTCCATATGTTGTTTCT 720
              |  || || |  |   ||| |  | |    |  ||  | |   ||||    | |
Db      16252 AACATAATTATAATAATTATAATATAATATTATTATAATATATAATATATAACATATAA 16193

Qy        721 GAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAAAGGAATTAAAAGCAATA 780
              | | |||      |  || | |     | |  | |  ||  | |  | ||| |   ||
Db      16192 TATTATATCTTATATAATATATAATATATAAGATATATATTATATAATATAATATATATT 16133

Qy        781 ATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTTT 820
              ||| |||  |||||    ||| |  ||  |   |||| ||
Db      16132 ATATATATTATACAATATATATAATATATAATATTATATT 16093


RESULT 8
US-11-098-686-8739/c
; Sequence 8739, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
;  APPLICANT: Kapur, Vivek and Gebhart, Connie J.
;  TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
;  TITLE OF INVENTION:  FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
;  FILE REFERENCE: 09531-128001
;  CURRENT APPLICATION NUMBER: US/11/098,686
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: PCT/US03/31318
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: US 60/416,395
;  PRIOR FILING DATE: 2002-10-04
;  NUMBER OF SEQ ID NOS: 11433
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8739
;   LENGTH: 1457619
;   TYPE: DNA
;   ORGANISM: Lawsonia intracellularis
US-11-098-686-8739

  Query Match             4.7%;  Score 55.2;  DB 17;  Length 1457619;
  Best Local Similarity   50.8%;  Pred. No. 19;
  Matches  132;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy        319 AAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGGTAAT 378
              | || || || |  | |   ||  || |  ||| | |||  || | |||| || ||
Db     801935 ATACATAAAAAATATAATAATATATAACATATTTTTATACATAAAAATTTTTTTAGTTAT 801876

Qy        379 ATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATC 438
              ||||      ||  | |  |||| | ||   | |  || |||| |  | |  | |     |
Db     801875 TTAGTAAAACTTAAAATTTTTATAATATAAATTTATAATTAATTATACTAATGTATATAC 801816

Qy        439 ATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTC 498
              |||  | | |||| |   || |||  | |||||  ||    || |  ||      ||  |
Db     801815 ATAAATTTTTATTATTATACTATTTATATATTTATTTTGTATTATTTATACTCATTTTTA 801756

Qy        499 ATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATA 558
              || || | |||| |   | ||  | |||    || |  || |  ||| | |  | |  |
Db     801755 TTTTTTTATTATAAAAGTAATAAAATTATTTATACATACTATTAATGTAGTTATTTATCA 801696

Qy        559 AAAAGAATAATAGGGTATTT 578
              | |  | ||| |   |  ||
Db     801695 ATATTATTAAGATATTGGTT 801676


RESULT 9
US-10-995-561-13313/c
; Sequence 13313, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13313
;   LENGTH: 26772
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13313

  Query Match             4.6%;  Score 54.8;  DB 10;  Length 26772;
  Best Local Similarity   43.1%;  Pred. No. 11;
  Matches  305;  Conservative    2;  Mismatches  399;  Indels    2;  Gaps    1;

Qy        340 AATGGACTTTACATTAGATATATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATT 399
              |||  |  ||| ||||  |||||  | || | |           |  | ||   |  |||
Db     21855 AATATATGTTATATTATTTATATAGAATACTATATAKTATATATAATATTATATTATATT 21796

Qy        400 ATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCATATTTGTATATTTTACAACG 459
              ||   |||| || |  |||| |  | |  ||      || ||| || |||||       |
Db     21795 ATATATTATAATTATAATAATTATATAATGATTTATATAATAATTATATATAAATATATA 21736

Qy        460 ATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCTTCAATATATTGATT 519
              ||| | |||||   ||   |  |||  | |  | | | |  | ||||||  |||
Db     21735 ATT--TATATTTATATAATTTATATAATAATTATATATAAATATATATAATATAAATT 21678

Qy        520 ACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTT 579
              |       |  |  |||      ||   | | || | | |||| ||        |
Db     21677 ATATTATATATCATATTGTATATAAAATATATAATATTATATATAATATTATATATAAAT 21618

Qy        580 TTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAA 639
              ||  |          |          |    | | ||  |     | |  |   | || |
Db     21617 TTTATATATATAATATATATTATATATAATATTATATATAATATATATTATATATAATAT 21558

Qy        640 AATGACTCATATAATAATATGTTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAA 699
              | | |   |||| ||          |   ||  || ||  |  | |  |     | ||
Db     21557 ATTTATATATATTATGTATAAGATATATTATATATTAATTATATATAATTAATAATTATG 21498

Qy        700 GGTGTTCCATATGTTGTTTCTGAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTA 759
              |    |||| || | | | |||    ||  |  |||  | |  |     |  |||
Db     21497 TATAAAATATATATTATATATTAATTATATATAATTGATAATTATGTATAAGATATATTA 21438

Qy        760 GAAAAGGAATTAAAAGCAATAATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTT 819
              | |   | |  | |  | |||||| ||      |   |   |      |  | |||| |
Db     21437 TATATTAATTATATATAATTAATAATTATGTATAAAATATATATTATATATTAATTATAT 21378

Qy        820 TATGGAGAACGGTTACATAAACAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTT 879
              ||  | ||      ||| |  ||||      | ||  |  | ||| |  | | | |  |
Db     21377 AATTAATAATTATATATAAAATATATTATTATATTATATATAAAAATATTATATATAAAAT 21318

Qy        880 ACTACGTATGGAGCAGAACTGTTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCT 939
              | ||  |||  |  |  | | ||  | | ||   | || |   |  |  ||  ||  |
Db     21317 ATTATATATTAAATAATATTATTTTATATGTTATATATTATACATTATTAGATATTCTAT 21258

Qy        940 TTAGGGATATATATACCTTTTTATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATA 999
              |      | ||||||  | | |||  ||  |     ||   | | |   |     | |||
Db     21257 ATTATATTTTATATAATTATATATAATTATATACACATTTTATATAGAAAATATTATATA 21198

Qy       1000 AATTGCGCATTCTATTCATATATCATTATGATTTTATTGCAATACTTA 1047
              | |       | || |  | |:|   ||  ||| |||   :|   |||
Db     21197 TAATATTATATATAATATTTTWTATATAAAATTATATATAWAATATTA 21150


RESULT 10
US-10-995-561-13479/c
; Sequence 13479, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
```

```
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13479
;    LENGTH: 54946
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-995-561-13479

  Query Match             4.6%;  Score 54.8;  DB 10;  Length 54946;
  Best Local Similarity   43.1%;  Pred. No. 12;
  Matches  305;  Conservative    2;  Mismatches  399;  Indels    2;  Gaps    1;

Qy        340 AATGGACTTTACATTAGATATATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATT 399
              |||  |  ||| ||||  |||||  | || | |          |  | ||    |  |||
Db       3309 AATATATGTTATATTATTTATATAGAATACTATATAKTATATATAATATTATATTATATT 3250

Qy        400 ATTCTTTATGATGAGCATAAATTCAAACAGAGGACCATCATATTTGTATATTTTACAACG 459
              ||   |||| || |  |||| |  | |  ||     || ||| || |||||       |
Db       3249 ATATATTATAATTATAATAATTATATAATGATTTATATAATAATTATATATAAATATATA 3190

Qy        460 ATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTTGCTCATTCTTCAATATATATTGATT 519
              |||  | |||||   ||   |    |||   | |   | | | |  | ||||||  |||
Db       3189 ATT--TATATTTATATAATTTATATAATAATTATATATAAATATATATAATATATAAATT 3132

Qy        520 ACCAATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTT 579
              |         |   | | |||      ||   | | || | | |||| ||        |
Db       3131 ATATTATATATCATATTGTATATAAAATATATAATATTATATATAATATTATATATAAAT 3072

Qy        580 TTATTGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAA 639
              ||  |        |     |    | | ||  |     | |  |   | || |
Db       3071 TTTATATATATAATATATATTATATATAAATATTATATATAATATATTATATATAATAT 3012

Qy        640 AATGACTCATATAATAATATGTTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAA 699
              | | |   |||| ||       |   ||  || ||  |  | |  |      | ||
Db       3011 ATTTATATATATTATGTATAAGATATATTATATATTAATTATATATAATTAATAATTATG 2952

Qy        700 GGTGTTCCATATGTTGTTTCTGAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTA 759
                |    |||| || | | | |||    ||  |  |||   | |      |      |||
Db       2951 TATAAAATATATATTATATATTAATTATATATAATTGATAATTATGTATAAGATATATTA 2892

Qy        760 GAAAAGGAATTAAAAGCAATAATAAATAGAATACAGGGAATAAAGCATCAAGACTTATTT 819
              | |   | |  | |  | |||||| ||      |   |  |     |  | |||| |
Db       2891 TATATTAATTATATATAATTAATAATTATGTATAAAATATATATTATATATTAATTATAT 2832

Qy        820 TATGGAGAACGGTTACATAAACAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTT 879
              ||  | ||   |     ||| |  ||||     | ||   |  | ||| | ||| |  |
Db       2831 AATTAATAATTATATATAAAATATATTATTATTATATATAAAAATATTATATATAAAAT 2772

Qy        880 ACTACGTATGGAGCAGAACTGTTAGTTTTTTTTGGTTTTCTCTGTGTATTCATTATCCCT 939
              | ||  |||  |  |  | | ||  | | ||   | || |   |  ||  ||  |
Db       2771 ATTATATATTAAATAATATTATTTTATATGTTATATATTATACATTATTAGATATTCTAT 2712

Qy        940 TTAGGGATATATATACCTTTTTATCTTTTAAAGAGAATGAAAAAAACCCATAGCTCGATA 999
              |   | ||||||  | | |||  ||  |     ||   | | |   |   |  |||
Db       2711 ATTATATTTTATATAATTATATATAATTATATACACATTTTATATAGAAAATATTATATA 2652

Qy       1000 AATTGCGCATTCTATTCATATATCATTATGATTTTATTGCAATACTTA 1047
              | |      | || |  | |:|   ||  ||| |||   :|   |||
Db       2651 TAATATTATATATAATATTTTWTATATAAAATTATATATAWAATATTA 2604


RESULT 11
US-11-011-332A-2
; Sequence 2, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
;  APPLICANT: Foekens, John
;  APPLICANT:  Harbeck, Nadia
;  APPLICANT:  Koenig, Thomas
;  APPLICANT:  Maier, Sabine
;  APPLICANT:  Martens, John
;  APPLICANT:  Model, Fabian
;  APPLICANT:  Nimmrich, Inko
;  APPLICANT:  Rujan, Tamas
;  APPLICANT:  Schmitt, Manfred
;  APPLICANT:  Lesche, Ralf
;  APPLICANT:  Dietrich, Dimo
;  APPLICANT:  Mueller, Volkmar
;  APPLICANT:  Kluth, Antje
;  APPLICANT:  Schwope, Ina
;  APPLICANT:  Hartmann, Oliver
;  APPLICANT:  Adorjan, Peter
;  TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SU
;  TITLE OF INVENTION:  BREAST CELL PROLIFERATIVE DISORDER PATIENTS
;  FILE REFERENCE: 47675-99
;  CURRENT APPLICATION NUMBER: US/11/011,332A
;  CURRENT FILING DATE:  2004-12-13
;  PRIOR APPLICATION NUMBER: US 10/517,741
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: PCT/EP2003/010881
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: DE 10245779.4
;  PRIOR FILING DATE: 2002-10-01
;  PRIOR APPLICATION NUMBER: DE 10300096.8
;  PRIOR FILING DATE: 2003-01-07
;  PRIOR APPLICATION NUMBER: DE 10317955.0
;  PRIOR FILING DATE: 2003-04-17
;  PRIOR APPLICATION NUMBER: PCT/EP2004/014170
;  PRIOR FILING DATE: 2004-12-13
;  PRIOR APPLICATION NUMBER: EP 03090432.0
;  PRIOR FILING DATE: 2003-12-11
;  PRIOR APPLICATION NUMBER: EP 04090041.7
;  PRIOR FILING DATE: 2004-02-10
;  PRIOR APPLICATION NUMBER: EP 04090380.9
;  PRIOR FILING DATE: 2004-09-30
;  PRIOR APPLICATION NUMBER: EP 04090127.4
;  PRIOR FILING DATE: 2004-04-01
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 2
;    LENGTH: 9001
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-2

  Query Match             4.6%;  Score 54.6;  DB 17;  Length 9001;
  Best Local Similarity   44.4%;  Pred. No. 9.3;
  Matches  265;  Conservative    0;  Mismatches  329;  Indels    3;  Gaps    1;

Qy         15 TAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCAT 74
              | ||||||  ||| |||   || ||| | |   |  | |    || |       |  ||
Db       7841 TTAGTTGAGGTTTATATATTTAGTTATATTAGGTTTTGTAATTTTATTATTAAATTATAT 7900

Qy         75 AGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCA 134
              | || ||  ||  |     || |  |  |||||   |   | |  |||  |  |  ||
Db       7901 ATATTATATTAGTAGTTTGTTGGTAAAGAAGGTTAAATTAATTTATATTTTGTTTATTAT 7960

Qy        135 GATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAA 194
              |  |||   |    |||||||| |       | |     || |             | |
Db       7961 TTGGTGTTTAAATGACGTATTTTATTTCGGAGATTTGGCGGAGAATTTTTTTTTTAGATT 8020

Qy        195 ATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAAT 254
              | || |||    | | || ||| |||  ||    | |     ||| | | |  | ||
Db       8021 TTATAGCGTTTTATTGAAGATAATGTTTTTATATTTGTAGTGGTTTTTAATTTGATAAGA 8080

Qy        255 TGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGG 314
              |  |  | |  ||   ||| ||||| ||||  ||   || |||   |  | |
Db       8081 TTTTAATTTGTTTAAGTTTTTTAAATAAGGGTTTTAAATGTTTTTAGTCGTTTTTTTATT 8140

Qy        315 GGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGG 374
              |  |    ||  | |   | ||| | ||  | | ||  |    || | ||| ||||||
Db       8141 GAATTTTTTTTAATTTTTTTAAGATTATAAAGTATATGT---GTAAAGTAAATATTTTTT 8197
```

```
Qy        375 TAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGAC 434
Db       8198 TTTATTGTATTGTTAGTCGATGATTTATAATTAAGTTAATAAGAATTTAGTTTTTTTTTG 8257

Qy        435 CATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTT 494
Db       8258 TTGAATGTGTTTATTAATTATATTTTAGTTTTTTTTTTTAAATTTTAGAATAGTTGTGGT 8317

Qy        495 GCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAA 554
Db       8318 TTTTATAATATTATGTTTTTTAAAGTTTTATTTTATGAAGGGATTTTATTATATTAAAGA 8377

Qy        555 AATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTA 611
Db       8378 ATGAAAAAAATTTTTATTGTAGTTAGTATATATAGTTTTTTATTTTTTGTTTTTTAA 8434
```

RESULT 12
US-11-011-332A-151
; Sequence 151, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
; APPLICANT: Foekens, John
; APPLICANT: Harbeck, Nadia
; APPLICANT: Koenig, Thomas
; APPLICANT: Maier, Sabine
; APPLICANT: Martens, John
; APPLICANT: Model, Fabian
; APPLICANT: Nimmrich, Inko
; APPLICANT: Rujan, Tamas
; APPLICANT: Schmitt, Manfred
; APPLICANT: Lesche, Ralf
; APPLICANT: Dietrich, Dimo
; APPLICANT: Mueller, Volkmar
; APPLICANT: Kluth, Antje
; APPLICANT: Schwope, Ina
; APPLICANT: Hartmann, Oliver
; APPLICANT: Adorjan, Peter
; TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SU
; TITLE OF INVENTION: BREAST CELL PROLIFERATIVE DISORDER PATIENTS
; FILE REFERENCE: 47675-99
; CURRENT APPLICATION NUMBER: US/11/011,332A
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 10/517,741
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: PCT/EP2003/010881
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: DE 10245779.4
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: DE 10300096.8
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: DE 10317955.0
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: PCT/EP2004/014170
; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: EP 03090432.0
; PRIOR FILING DATE: 2003-12-11
; PRIOR APPLICATION NUMBER: EP 04090041.7
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: EP 04090380.9
; PRIOR FILING DATE: 2004-09-30
; PRIOR APPLICATION NUMBER: EP 04090127.4
; PRIOR FILING DATE: 2004-04-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 151
; LENGTH: 28536
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-151

Query Match 4.6%; Score 54.6; DB 17; Length 28536;
Best Local Similarity 44.4%; Pred. No. 12;
Matches 265; Conservative 0; Mismatches 329; Indels 3; Gaps 1;

```
Qy         15 TAAGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCAT 74
Db       9514 TTAGTTGAGGTTTATATATTTAGTTATATTAGGTTTTGTAATTTTATTATTAAATTATAT 9573

Qy         75 AGATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCA 134
Db       9574 ATATTATATTAGTAGTTTGTTGGTAAAGAAGGTTAAATTAATTTATATTTGTTTATTAT 9633

Qy        135 GATTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAA 194
Db       9634 TTGGTGTTTAAATGACGTATTTTATTTCGGAGATTTGGCGGAGAATTTTTTTTTTAGATT 9693

Qy        195 ATTTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAAT 254
Db       9694 TTATAGCGTTTTATTGAAGATAATGTTTTTATATTTGTAGTGGTTTTTAATTTGATAAGA 9753

Qy        255 TGCTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGATCCAGGGGG 314
Db       9754 TTTTAATTTGTTTAAGTTTTTTAAATAAGGGTTTTAAATGTTTTTAGTCGTTTTTTTATT 9813

Qy        315 GGCTAAACTTAGAATAGCTGAAGGTAATGGACTTTACATTAGATATATTAAGTATTTTGG 374
Db       9814 GAATTTTTTTTAATTTTTTTAAGATTATAAAGTATATGT---GTAAAGTAAATATTTTTT 9870

Qy        375 TAATATAGTTGTGTTTGCATTAATTATTCTTTATGATGAGCATAAATTCAAACAGAGGAC 434
Db       9871 TTTATTGTATTGTTAGTCGATGATTTATAATTAAGTTAATAAGAATTTAGTTTTTTTTTG 9930

Qy        435 CATCATATTTGTATATTTTACAACGATTGCTTTATTTGGTTATCGTTCTGAATTGGTGTT 494
Db       9931 TTGAATGTGTTTATTAATTATATTTTAGTTTTTTTTTTTAAATTTTAGAATAGTTGTGGT 9990

Qy        495 GCTCATTCTTCAATATATATTGATTACCAATATCCTGTCAAAGGATAACCGTAATCCTAA 554
Db       9991 TTTTATAATATTATGTTTTTTAAAGTTTTATTTTATGAAGGGATTTTATTATATTAAAGA 10050

Qy        555 AATAAAAAGAATAATAGGGTATTTTTTATTGGTAGGGGTTGTATGCTCGTTGTTTTA 611
Db      10051 ATGAAAAAAATTTTTATTGTAGTTAGTATATATAGTTTTTTATTTTTTGTTTTTTAA 10107
```

RESULT 13
US-09-925-065A-436829
; Sequence 436829, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 436829
; LENGTH: 596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-436829

```
  Query Match             4.5%;  Score 53.2;  DB 7;  Length 596;
  Best Local Similarity   49.0%;  Pred. No. 9.7;
  Matches  142;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy         17 AGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATAG 76
              | ||  ||  |||  |    ||||    | ||||| | || || ||||| || |    |
Db        104 AATTTTTTCCTTTTCTGTGTTTTAATTCTGATAGTTTGCATTTTTATGGTGTGAGAGCAA 163

Qy         77 ATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGA 136
              ||  |   |    |   | |||  |     || ||  |   | ||   || |||||  |
Db        164 ATTTTGAAAAGGTACTCTATGAAGCCTCCATATRTATTTTTCAATAATTTTTTATTTGATG 223

Qy        137 TTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAAT 196
               ||  |||  |  |||||||  |  |||  | |  |||  |||  |   ||  || |    |
Db        224 GTTGCTTTTCTGCATTATTTCAAATAAAAAATGAAAAATAAAAATTTCAAAAGTAGCTTT 283

Qy        197 TTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTG 256
               ||  | |  |  ||  ||    |||||     |||   ||||       || |  |
Db        284 GGATATTAACTTTTGATATATTAGTAATAAAATATTTTTATTTTACATAATTAAATTTGC 343

Qy        257 CTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGAT 306
              || ||||  ||  ||  || ||  ||  |       |  ||  | |||||
Db        344 CTCCTGTTTTTCAATAATTTGAACCAGACAGGTTGTTTATTTGTTATGAT 393


RESULT 14
US-10-301-480-498452
; Sequence 498452, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 498452
;   LENGTH: 597
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-498452

  Query Match             4.5%;  Score 53.2;  DB 12;  Length 597;
  Best Local Similarity   49.0%;  Pred. No. 9.7;
  Matches  142;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy         17 AGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATAG 76
              | ||  ||  |||  |    ||||    | ||||| | || || ||||| || |    |
Db        105 AATTTTTTCCTTTTCTGTGTTTTAATTCTGATAGTTTGCATTTTTATGGTGTGAGAGCAA 164

Qy         77 ATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGA 136
              ||  |   |    |   | |||  |     || ||  |   | ||   || |||||  |
Db        165 ATTTTGAAAAGGTACTCTATGAAGCCTCCATATRTATTTTTCAATAATTTTTTATTTGATG 224

Qy        137 TTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAAT 196
               ||  |||  |  |||||||  |  |||  | |  |||  |||  |   ||  || |    |
Db        225 GTTGCTTTTCTGCATTATTTCAAATAAAAAATGAAAAATAAAAATTTCAAAAGTAGCTTT 284

Qy        197 TTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTG 256
               ||  | |  |  ||  ||    |||||     |||   ||||       || |  |
Db        285 GGATATTAACTTTTGATATATTAGTAATAAAATATTTTTATTTTACATAATTAAATTTGC 344

Qy        257 CTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGAT 306
              || ||||  ||  ||  || ||  ||  |       |  ||  | |||||
Db        345 CTCCTGTTTTTCAATAATTTGAACCAGACAGGTTGTTTATTTGTTATGAT 394


RESULT 15
US-10-301-480-1111861
; Sequence 1111861, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1111861
;   LENGTH: 597
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1111861

  Query Match             4.5%;  Score 53.2;  DB 12;  Length 597;
  Best Local Similarity   49.0%;  Pred. No. 9.7;
  Matches  142;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy         17 AGTTGATTTTTTTATTCCTATTTACACTTTATAGTCTCCAGTTGTATGGGGTTATCATAG 76
              | ||  ||  |||  |    ||||    | ||||| | || || ||||| || |    |
Db        105 AATTTTTTCCTTTTCTGTGTTTTAATTCTGATAGTTTGCATTTTTATGGTGTGAGAGCAA 164

Qy         77 ATGATCGTATAACAAATTTTGATACAAAGGTATTAACTAGTATTATAATTATATTTCAGA 136
              ||  |   |    |   | |||  |     || ||  |   | ||   || |||||  |
Db        165 ATTTTGAAAAGGTACTCTATGAAGCCTCCATATRTATTTTTCAATAATTTTTTATTTGATG 224

Qy        137 TTTTTTTTGTTTTATTATTTTATCTAACGATTATAAATGAAAGAAAACAGCAGAAAAAAT 196
               ||  |||  |  |||||||  |  |||  | |  |||  |||  |   ||  || |    |
Db        225 GTTGCTTTTCTGCATTATTTCAAATAAAAAATGAAAAATAAAAATTTCAAAAGTAGCTTT 284

Qy        197 TTATCGTGAACTGGGAGCTAAAGTTAATACTCGTTTTCCTTTTTGTGACTATAGAAATTG 256
               ||  | |  |  ||  ||    |||||     |||   ||||       || |  |
Db        285 GGATATTAACTTTTGATATATTAGTAATAAAATATTTTTATTTTACATAATTAAATTTGC 344

Qy        257 CTGCTGTAGTTTTATTTCTTAAAGAAGGTATTCCTATATTTGATGATGAT 306
              || ||||  ||  ||  || ||  ||  |       |  ||  | |||||
Db        345 CTCCTGTTTTTCAATAATTTGAACCAGACAGGTTGTTTATTTGTTATGAT 394


Search completed: May 11, 2006, 08:30:31
Job time : 616.35 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 08:31:04 ; Search time 3466.17 Seconds
(without alignments)
15792.702 Million cell updates/sec

Title: US-09-423-093B-3_COPY_2352_3314
Perfect score: 963
Sequence: 1 atgagtaaaaaacctcttct.....cggttttaatgcggaaataa 963

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	963	100.0	12441	1	SERFBC2	X61917 S.enterica
2	963	100.0	12441	6	BD085556	BD085556 Nucleic a
3	963	100.0	12446	6	AX029263	AX029263 Sequence
4	73.8	7.7	110000	1	BA000019_34	Continuation (35 o
5	63.2	6.6	110000	1	AE017197_05	Continuation (6 of
6	62.8	6.5	110000	1	AE017308_5	Continuation (6 of
7	61.6	6.4	15000	1	AY253301	AY253301 Actinobac
8	60.8	6.3	19457	1	AF461769	AF461769 Yersinia
9	60.8	6.3	20125	1	AF461770	AF461770 Yersinia
10	59	6.1	312430	1	RPXX02	AJ235271 Rickettsi
11	58.8	6.1	196988	8	AP002761	AP002761 Homo sapi
12	57.6	6.0	26913	1	YEU46859	U46859 Yersinia en
13	57	5.9	20511	1	AF461768	AF461768 Yersinia
14	56.4	5.9	106063	5	AL929058	AL929058 Zebrafish
15	56	5.8	110000	1	CP000020_01	Continuation (2 of
16	56	5.8	110000	1	CP000020_02	Continuation (3 of
17	56	5.8	250022	2	AE014824	AE014824 Plasmodiu
18	55.8	5.8	11186	1	AE013710	AE013710 Yersinia
19	55.8	5.8	20588	1	YPS251712	AJ251712 Yersinia
20	55.8	5.8	20745	1	YPE251713	AJ251713 Yersinia
21	55.8	5.8	76121	8	AC008463	AC008463 Homo sapi
22	55.8	5.8	110000	1	BX936398_11	Continuation (12 o
23	55.8	5.8	110000	1	BX936398_12	Continuation (13 o
24	55.8	5.8	140502	9	AL773552	AL773552 Mouse DNA
c 25	55.8	5.8	205050	1	AJ414155	AJ414155 Yersinia
26	55.8	5.8	290510	1	AE017130	AE017130 Yersinia
c 27	55.4	5.8	158548	2	PFMAL3P2	AL034558 Plasmodiu
28	55.2	5.7	263051	14	CR450811	CR450811 Danio rer
c 29	55	5.7	250663	2	AE014826	AE014826 Plasmodiu
c 30	54.8	5.7	167954	1	BX842645	BX842645 Mycoplasm
31	54.8	5.7	255008	14	BX510915	BX510915 Danio rer
c 32	54.8	5.7	347558	1	BX842642	BX842642 Mycoplasm
c 33	54.6	5.7	12029	2	AE001428	AE001428 Plasmodiu
c 34	54.6	5.7	192929	14	AC005505	AC005505 Plasmodiu
35	54.6	5.7	209342	5	BX470150	BX470150 Zebrafish
c 36	54.6	5.7	250713	2	AE014850	AE014850 Plasmodiu
37	54.6	5.7	256172	14	AC005139	AC005139 Plasmodiu
38	54.4	5.6	1141	6	AR579680	AR579680 Sequence
39	54.4	5.6	1141	6	AX083744	AX083744 Sequence
40	54.4	5.6	110000	14	PFMAL13_06	Continuation (7 of
c 41	54.4	5.6	148726	5	BX005402	BX005402 Zebrafish
c 42	54.4	5.6	165319	5	BX682234	BX682234 Zebrafish
43	54.4	5.6	283619	14	CR753823	CR753823 Danio rer
c 44	54.2	5.6	110000	14	PFMAL13_02	Continuation (3 of
c 45	54	5.6	55372	14	CR759965	CR759965 Danio rer

ALIGNMENTS

RESULT 1
SERFBC2
LOCUS SERFBC2 12441 bp DNA linear BCT 09-SEP-2004
DEFINITION S.enterica rfbJ gene cluster.
ACCESSION X61917
VERSION X61917.1 GI:47004
KEYWORDS abe gene; abequose synthase; galactosyltransferase; manB gene; manC gene; mannose-1-phosphate guanyltransferase; phosphomannomutase; wbaL gene; wbaP gene; wbaQ gene; wbaR gene; wbaW gene; wbaZ gene; wzx gene; wzy gene.
SOURCE Salmonella enterica
ORGANISM Salmonella enterica
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 12441)
AUTHORS Brown,P.K., Romana,L.K. and Reeves,P.R.
TITLE Molecular analysis of the rfb gene cluster of Salmonella serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B
JOURNAL Mol. Microbiol. 6 (10), 1385-1394 (1992)
PUBMED 1379320
REFERENCE 2
AUTHORS Reeves,P.
TITLE Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale
JOURNAL Trends Genet. 9 (1), 17-22 (1993)
PUBMED 8434412
REFERENCE 3
AUTHORS Liu,D., Haase,A.M., Lindqvist,L., Lindberg,A.A. and Reeves,P.R.
TITLE Glycosyl transferases of O-antigen biosynthesis in Salmonella enterica: identification and characterization of transferase genes of groups B, C2, and E1
JOURNAL J. Bacteriol. 175 (11), 3408-3413 (1993)
PUBMED 7684736
REFERENCE 4 (bases 1 to 12441)
AUTHORS Brown,P.K.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1991) P.K. Brown, University of Sydney, Dept of Microbiology, Maze Crescent, Sydney N S W 2006, AUSTRALIA
FEATURES Location/Qualifiers
source 1. .12441

/organism="Salmonella enterica"
/mol_type="genomic DNA"
/strain="group C2, serovar muenchen M67"
/db_xref="taxon:28901"
gene 138. .1019
/gene="abe"
CDS 138. .1019
/gene="abe"
/codon_start=1
/transl_table=11
/product="abequose synthase"
/protein_id="CAA43918.1"
/db_xref="GI:47005"
/db_xref="GOA:Q00329"
/db_xref="UniProt/Swiss-Prot:Q00329"
/translation="MLDVNKKILMTGATSFVGTHLLHSLIKEGYSIIALKRPITEPTI
INTLIEWLNIQDIEKICQSSMNIHAIVHIATDYGRNRTPISEQYKCNVLLPTRLLELM
PALKTKFFISTDSFFGKYEKHYGYMRSYMASKRHFVELSKIYVEEHPDVCFINLRLEH
VYGERDKAGKIIPYVIKKMKNNEDIDCTIARQKRDFIYIDDVVSAYLKILKEGFNAGH
YDVEVGTGKSIELKEVFEIIKKETHSSSKINYGAVAMRDDEIMESHANTSFLTRLGWS
AEFSIEKGVKKMLSMKE"
CDS 1019. .2359
/note="wzx
Referred to as orf12.6 in reference [2]"
/citation=[2]
/codon_start=1
/transl_table=11
/product="hypothetical protein with 12 predicted
transmembrane segments which possibly functions in export
of O-antigen"
/protein_id="CAA43919.1"
/db_xref="GI:47006"
/db_xref="UniProt/TrEMBL:Q00339"
/translation="MNRIIRMLGVDKAIRYVIFGKIISVLTGLLLIMLISHHLSKDAQ
GYYYTFNSVVALQIIFELGLSTVIIQFASHEMSALKYDYSERDIIGESKNKQRYLSLF
RLAIKWYAVIALLIILIVGPIGYVFFTQKEGLGVPWQGAWLLLTIVTAFNIFLVSVLS
VAEGSGLITDVNKMRMYQSLLAGILAVSLLISGFGLYATSAIAISGTIIFSIFSYKYF
KKIFLQSLKHKNKYTEGGISWVNEIFPMQWRIALSWMSGYFIYFVMTPIAFKYFGAIY
AGQLGMSLTLCNMVMATGLAWISTKYPKWGVMVSNKQLAELSKSFKSAVMQSSFFVLT
GLTGVYISLWLLKLSGSNIGERFLGLQDFFFLSLAIIGNHIVACFATYIRAHKTEKMT
LASCIMALLTITTMLFVAYLEYSRFYMLMYAALTWLYFVPQTYIIFKRFKSSYE"
CDS 2352. .3314
/note="unnamed protein product; wbaR"
/codon_start=1
/transl_table=11
/protein_id="CAA43909.1"
/db_xref="GI:47007"
/db_xref="UniProt/TrEMBL:Q00337"
/translation="MSKKPLLTIAIPTYNRSSCLARLLDSIIQQENYCHDELEVIVCD
NASTDETARIAKSGLDKIRNSTYHLNEENLGMDGNFQKCFELSNGKYLWMIGDDDLIV
KNGISKVFSILKSRPALDMVYVNSAAKTELNYNADVRTSFYTNDVDFISDVKVMFTFI
SGMICKKTDAIVKAVGIFSPQTTGKYLMHLTWQLPLLKQGGEFAVIHNNIIEAEPDNS
GGYHLYKVFSNNLATIFDVFYPREHRVSKRVRASACLFLLNFIGDEDKTKNFATNNYL
RDCDSAFIDLIIYKYGLRFFYLYPKTVPLFRKIKYIIKTVLMRK"
CDS 3361. .3975
/note="unnamed protein product; wbaL
orf14.9 possibly encodes an acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="CAA43910.1"
/db_xref="GI:47008"
/db_xref="UniProt/TrEMBL:Q00480"
/translation="MFVYSLRLKLNLIISLLSKVRRKSKAKFLVLLSGYDFKMVGKNF
KLNVKPYSAKNNTSSKWGSMRVGDNCWIEAVYNYGDEKFEPYLYIGDRICLSDNVHIS
CVSCLILENDILIGSKVYIGDHSHGSYKVCSPKIEPPANKPLGDIAPIKIGNCCWIGD
NAVILAGSEICDGCVIAANSVVKDLKVDKPCLIGGVPAKVIKVF"
CDS 3977. .5020
/note="wbaQ
Referred to as orf15.6 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="rhamnosyl transferase"
/protein_id="CAA43911.1"
/db_xref="GI:47009"
/db_xref="GOA:Q00482"
/db_xref="UniProt/TrEMBL:Q00482"
/translation="MNVFISICIPSYNRAEFLEPLLDSIYNQDYCLKNNDFEVIVCED
KSPQRDEINSIIENYKAKNNKQNLYVNFNEDNLGYDKNLKKCISLTTGKYCMIMGNDD
LLADGALSKIVKVLKANPEIVLATRAYGWFKENPNELCDTVRHLTDDTLFQPGADAIK
FFFRRVGVISGFIVNAEKAKKLSSDLFDGRLYYQMYLAGMLMAEGQGYYFSDVMTLSR
DTEAPDFGNAGTEKGVFTPGGYKPEGRIHMVEGLLLIAKYIEDTTKIDGVYAGIRKDL
ANYFYPYIRDQLDLPLYTYIKMINKFRKMGFSNEKLFYVHAFLGYVLKRRGYDALIKY
IRSKKGGTPRLGI"
gene 5114. .6313
/gene="wzy"
CDS 5114. .6313
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O antigen polymerase"
/protein_id="CAA43912.1"
/db_xref="GI:47010"
/db_xref="GOA:Q00474"
/db_xref="UniProt/Swiss-Prot:Q00474"
/translation="MLPFPPGAILRDVLNVFFVALVLVRFVIDRKKTYFPLVFTIFSW
SAVILWVIALTIFSPDKIQAIMGGRSYILFPAVFIALVILKVSYPQSLNIEKIVCYII
FLMFMVATISIIDVLMNGEFIKLLGYDEHYAGEQLNLINSYDGMVRATGGFSDALNFG
YMLTLGVLLCMECFSQGYKRLLMLIISFVLFIAICMSLTRGAILVAALIYALYIISNR
KMLFCGITLFVIIPVLAISTNIFDNYTEILIGRFTDSSQASRGSTQGRIDMAINSLN
FLSEHPSGIGLGTQGSGNMLSVKDNRLNTDNYFFWIALETGIIGLIINIIYLASQFYS
STLLNRIYGSHCSNMHYRLYFLFGSIYFISAALSSAPSSSTFSIYYWTVLALIPFLKL
TNRRCTR"
CDS 6313. .7323
/note="wbaW
Referred to as orf17.9 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="second mannosyl transferase"
/protein_id="CAA43913.1"
/db_xref="GI:47011"
/db_xref="GOA:Q00481"
/db_xref="UniProt/TrEMBL:Q00481"
/translation="MNNKKVLMDISWSNKGGIGRFTDEISKLLCDISKEELYRKCASP
LAPLGLAVNIFLRKKTDVVFLPGYIPPLFCSKKPIITIHDLNHLDLNDNSSLFKRLFY
NFIIKRGCRKAYKIFTVSNFSKERIVAWSGVNPNKIVTVYNGVSSLFNADVKPLNLGY
KYLLCVGNRKTHKNEKCVISAFAKADIDPSIKLVFTGNPCNDLEKLIIQHGLSERVKF
FGFVSEKDLPSLYKGSLGLVFPSLYEGFGLPVVEGMACGIPVLTSLTSSLPEVAGDAA
ILVDPLSEDAITKGISRLINDSELRKHLIQKGLLRAKRFNWQNVVSEIEMVLTEACDG
NK"
CDS 7310. .8467
/note="wbaZ
Referred to as orf18.9 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="first mannosyl transferase"
/protein_id="CAA43914.1"
/db_xref="GI:47012"
/db_xref="GOA:Q00483"
/db_xref="UniProt/TrEMBL:Q00483"
/translation="MEINEIKISLVHEWLLSYAGSEQVSSAILHVFPEAKLYSVVDFL
TDEQRRHFLGKYATTTFIQNLPKAKKFYQKYLPLMPLAIEQLDLSDANIIISSAHSVA
KGVISGPDQLHISYVHSPIRYAWDLQHQYLNESNLNKGIKGWLAKWLLHKIRIWDSRT
ANGVDHFIANSQYIARRIKKVYRREASVIYPPVDVDNFEVKNEKQDYYFTASRMVPYK
RIDLIVEAFSKMPEKKLVVIGDGPEMKKIKSKATDNIKLLGYQSFPVLKEYMQSARAF
VFAAEEDFGIIPVEAQACGTPVIAFGKGGALETVRPLGVEEPTGIFFKEQNIASLHEA
VSEFEKNASFFTSQACRKNAEKFSRSRFEQEFKNFVNEKWNLFKTEQIIKR"
gene 8482. .9903
/gene="manC"
CDS 8482. .9903
/gene="manC"
/codon_start=1
/transl_table=11
/product="mannose-1-phosphate guanyltransferase"

```
                     /protein_id="CAA43915.1"
                     /db_xref="GI:47013"
                     /db_xref="GOA:Q00473"
                     /db_xref="UniProt/Swiss-Prot:Q00473"
                     /translation="MSKLIPVIMAGGIGSRLWPLSREEHPKQFLSVDGELSMLQNTIK
                     RLTPLLAGEPLVICNDSHRFLVAEQLRAINKLANNIILEPVGRNTAPAIALAAFCSLQ
                     NVVDEDPLLLVLAADHVIRDEKVFLKAINHAEFFATQGKLVTFGIVPTQAETGYGYIC
                     RGEAIGEDAFSVAEFVEKPDFDTARHYVESEKYYWNSGMFLFRASSYLQELKDLSPDI
                     YQACENAVGSINPDLDFIRIDKEAFAMCPSDSIDYAVMEHTRHAVVVPMNAGWSDVGS
                     WSSLWDISKKDPQRNVLHGDIFAYNSKDNYIYSEKSFISTIGVNNLVIVQTADALLVS
                     DKDSVQDVKKVVDYLKANNRNEHKKHLEVFRPWGKFSVIHSGDNYLVKRITVKPGAKF
                     AAQMHLHRAEHWIVVSGTACITKGEEIFTISENESTFIPANTVHTLKNPATIPLELIE
                     IQSGTYLAEDDIIRLEKHSGYLE"
     gene            9911. .11347
                     /gene="manB"

  Query Match           100.0%;  Score 963;  DB 1;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 7.9e-146;
  Matches  963;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2352 ATGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTG 2411

Qy         61 GCTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2412 GCTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTT 2471

Qy        121 ATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2472 ATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGAT 2531

Qy        181 AAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2532 AAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTC 2591

Qy        241 CAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2592 CAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTA 2651

Qy        301 ATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2652 ATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGAT 2711

Qy        361 ATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2712 ATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACG 2771

Qy        421 TCATTCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2772 TCATTCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATT 2831

Qy        481 TCTGGAATGATATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2832 TCTGGAATGATATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCG 2891

Qy        541 CAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2892 CAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGT 2951

Qy        601 GGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2952 GGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGA 3011

Qy        661 TATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3012 TATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCC 3071

Qy        721 AGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3072 AGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTC 3131

Qy        781 ATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3132 ATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGAT 3191

Qy        841 AGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3192 AGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCT 3251

Qy        901 AAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAATGCGGAAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3252 AAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAATGCGGAAA 3311

Qy        961 TAA 963
              |||
Db       3312 TAA 3314


RESULT 2
BD085556
LOCUS       BD085556               12441 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085556
VERSION     BD085556.1  GI:22631166
KEYWORDS    JP 2001522248-A/3.
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 12441)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 3 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Salmonella enterica
            PN   JP 2001522248-A/3
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .12441
            FT                   /organism='Salmonella enterica'.
FEATURES             Location/Qualifiers
     source          1. .12441
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match           100.0%;  Score 963;  DB 6;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 7.9e-146;
  Matches  963;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2352 ATGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTG 2411

Qy         61 GCTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2412 GCTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTT 2471

Qy        121 ATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2472 ATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGAT 2531

Qy        181 AAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2532 AAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTC 2591

Qy        241 CAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTA 300
```

```
Db       2592 CAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTA 2651

Qy        301 ATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2652 ATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGAT 2711

Qy        361 ATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2712 ATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACG 2771

Qy        421 TCATTCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2772 TCATTCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATT 2831

Qy        481 TCTGGAATGATATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2832 TCTGGAATGATATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCG 2891

Qy        541 CAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2892 CAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGT 2951

Qy        601 GGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2952 GGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGA 3011

Qy        661 TATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3012 TATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCC 3071

Qy        721 AGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3072 AGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTC 3131

Qy        781 ATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3132 ATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGAT 3191

Qy        841 AGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3192 AGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCT 3251

Qy        901 AAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAATGCGGAAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3252 AAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAATGCGGAAA 3311

Qy        961 TAA 963
              |||
Db       3312 TAA 3314
```

```
RESULT 3
AX029263
LOCUS       AX029263               12446 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 3 from Patent WO9850531.
ACCESSION   AX029263
VERSION     AX029263.1  GI:10190114
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 3 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .12446
                     /organism="Salmonella enterica"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:28901"
ORIGIN
```

```
  Query Match           100.0%;  Score 963;  DB 6;  Length 12446;
  Best Local Similarity  100.0%;  Pred. No. 7.9e-146;
  Matches  963;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2352 ATGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTG 2411

Qy         61 GCTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2412 GCTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTT 2471

Qy        121 ATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2472 ATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGAT 2531

Qy        181 AAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2532 AAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTC 2591

Qy        241 CAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2592 CAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTA 2651

Qy        301 ATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2652 ATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGAT 2711

Qy        361 ATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2712 ATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACG 2771

Qy        421 TCATTCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2772 TCATTCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATT 2831

Qy        481 TCTGGAATGATATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2832 TCTGGAATGATATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCG 2891

Qy        541 CAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2892 CAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGT 2951

Qy        601 GGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2952 GGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGA 3011

Qy        661 TATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3012 TATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCC 3071

Qy        721 AGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3072 AGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTC 3131

Qy        781 ATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3132 ATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGAT 3191

Qy        841 AGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3192 AGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCT 3251

Qy        901 AAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAATGCGGAAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3252 AAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAATGCGGAAA 3311

Qy        961 TAA 963
              |||
```

```
Db       3312 TAA 3314


RESULT 4
BA000019_34
WPCOMMENT
  Sequence split into 65 fragments  LOCUS BA000019 Accession BA000019
     Fragment Name      Begin     End
     BA000019_00            1    110000
     BA000019_01       100001    210000
     BA000019_02       200001    310000
     BA000019_03       300001    410000
     BA000019_04       400001    510000
     BA000019_05       500001    610000
     BA000019_06       600001    710000
     BA000019_07       700001    810000
     BA000019_08       800001    910000
     BA000019_09       900001   1010000
     BA000019_10      1000001   1110000
     BA000019_11      1100001   1210000
     BA000019_12      1200001   1310000
     BA000019_13      1300001   1410000
     BA000019_14      1400001   1510000
     BA000019_15      1500001   1610000
     BA000019_16      1600001   1710000
     BA000019_17      1700001   1810000
     BA000019_18      1800001   1910000
     BA000019_19      1900001   2010000
     BA000019_20      2000001   2110000
     BA000019_21      2100001   2210000
     BA000019_22      2200001   2310000
     BA000019_23      2300001   2410000
     BA000019_24      2400001   2510000
     BA000019_25      2500001   2610000
     BA000019_26      2600001   2710000
     BA000019_27      2700001   2810000
     BA000019_28      2800001   2910000
     BA000019_29      2900001   3010000
     BA000019_30      3000001   3110000
     BA000019_31      3100001   3210000
     BA000019_32      3200001   3310000
     BA000019_33      3300001   3410000
     BA000019_34      3400001   3510000
     BA000019_35      3500001   3610000
     BA000019_36      3600001   3710000
     BA000019_37      3700001   3810000
     BA000019_38      3800001   3910000
     BA000019_39      3900001   4010000
     BA000019_40      4000001   4110000
     BA000019_41      4100001   4210000
     BA000019_42      4200001   4310000
     BA000019_43      4300001   4410000
     BA000019_44      4400001   4510000
     BA000019_45      4500001   4610000
     BA000019_46      4600001   4710000
     BA000019_47      4700001   4810000
     BA000019_48      4800001   4910000
     BA000019_49      4900001   5010000
     BA000019_50      5000001   5110000
     BA000019_51      5100001   5210000
     BA000019_52      5200001   5310000
     BA000019_53      5300001   5410000
     BA000019_54      5400001   5510000
     BA000019_55      5500001   5610000
     BA000019_56      5600001   5710000
     BA000019_57      5700001   5810000
     BA000019_58      5800001   5910000
     BA000019_59      5900001   6010000
     BA000019_60      6000001   6110000
     BA000019_61      6100001   6210000
     BA000019_62      6200001   6310000
     BA000019_63      6300001   6410000
     BA000019_64      6400001   6413771
```

Continuation (35 of 65) of BA000019 from base 3400001 (BA000019 Nostoc sp. PCC 7120 DNA,

```
  Query Match             7.7%;  Score 73.8;  DB 1;  Length 110000;
  Best Local Similarity   48.8%;  Pred. No. 0.0033;
  Matches  232;  Conservative    0;  Mismatches  237;  Indels    6;  Gaps    1;

Qy          2 TGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTGG 61
              |||  |  ||     | || || |||||| | || ||||||||| ||  ||    |   |||
Db      44423 TGATCATGAATAAATTACTCACAATTGCTCTCCCCACATATAATCGAGCTAAGTTACTGG 44482

Qy         62 CTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTTA 121
               ||    | | |   |  ||  | ||       ||   || | |   ||||   ||  | |
Db      44483 ATCAACAATTAGCATGGCTAGCTAAATCCATCAAAGGTTTTGAATCTGAATGTGAAATCA 44542

Qy        122 TTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGATA 181
              || ||| ||||||   ||||  |||| | |||  |  |||  |  || |   |
Db      44543 TTATTTCTGATAACTGTTCAGAAGATAACACACAAGAAATTGTGAAAAAATGGCAGATGG 44602

Qy        182 AAATAAGAAATAGTACTTATCATCT------AAATGAAGAAAACTTAGGAATGGATGGTA 235
              |   | | |  |||| |||          |||      ||||  | ||| ||    | |
Db      44603 CTTTGCCAGAAACAACTTTTCACTCCCACAGAAACAGTCAAAATCTGGGAGTGATGAGAA 44662

Qy        236 ACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATG 295
              |  |      | ||| || |     ||| | |  ||||| ||||||  ||  |||||||||
Db      44663 ATATTGCTTATTGTCTTAATGCTGCAACTAGCCAATATGTTTGGACAATCAGCGATGATG 44722

Qy        296 ATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCAT 355
              ||  ||| |   |       |||||  |  |  |    | |||||| |       |   |
Db      44723 ATAAAATCGAAGATGCAACGATTTCTTACCTCATCAACACATTAAAGACATCAGTTAATT 44782

Qy        356 TAGATATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGA 415
              ||| | || ||  | ||||||   |          ||| | |    | |
Db      44783 TAGGTTTGCTGATTTTAAATTTTTCTTGTCGCCATGAAGTGACGGGTGAGTTATTGTACC 44842

Qy        416 GGACGTCATTCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTT 470
              |  |  | | | |   | |||| || | ||   || || | || ||   | ||
Db      44843 AAAGCTGCTACCAAATTGAAGATGAAGTTGTTGAGTCTGATGGGAGAGCAGTTTT 44897
```

```
RESULT 5
AE017197_05
WPCOMMENT
  Sequence split into 12 fragments  LOCUS AE017197 Accession AE017197
     Fragment Name      Begin     End
     AE017197_00            1    110000
     AE017197_01       100001    210000
     AE017197_02       200001    310000
     AE017197_03       300001    410000
     AE017197_04       400001    510000
     AE017197_05       500001    610000
     AE017197_06       600001    710000
     AE017197_07       700001    810000
     AE017197_08       800001    910000
     AE017197_09       900001   1010000
     AE017197_10      1000001   1110000
     AE017197_11      1100001   1111496
Continuation (6 of 12) of AE017197 from base 500001 (AE017197 Rickettsia typhi str. Wilma

  Query Match             6.6%;  Score 63.2;  DB 1;  Length 110000;
  Best Local Similarity   50.0%;  Pred. No. 0.17;
  Matches  158;  Conservative    0;  Mismatches  158;  Indels    0;  Gaps    0;

Qy         94 AACTATTGTCATGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACA 153
              ||  |  ||  | | ||  | ||  ||||| ||  |||| ||  ||||| |||| | |
Db      46820 AATCAAAGTTTTAAAGATTTTGAACTTATTATTATTGATGATTGTTCAAAAGATAATAGT 46879

Qy        154 GCAAGAATAGCCAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAA 213
              || |  ||||| ||||        ||| |      |  || ||          | ||||||
Db      46880 GCTATGATAGCTAAGAAATATGCAGAGCAGGATTCACGTATTAAGATAATAATCAATGAA 46939

Qy        214 GAAAACTTAGGAATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATAT 273
              | ||      || |    |  |  ||  | |    |||| |     | ||||||||||||
Db      46940 CATAATAAGAGATTATCAGCAAGTTTGAATATTGCTTTTAAAAAGGCTAATGGAAAATAT 46999
```

```
Qy        274 CTTTGGATGATTGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCG 333
              ||      || |   |||||| |||| |     |  ||| | |  | ||  |  ||  |
Db      47000 TTTACTTGGACTTCAGATGATAATCTTTTCCATGAGGATGCTTTAGCAAAAATGGTTAAG 47059

Qy        334 ATATTAAAGTCCCGGCCTGCATTAGATATGGTGTATGTAAATTCAGCAGCAAAGACTGAG 393
              |||  |  |       ||||   |||   | || |||    |||   |   ||    ||||
Db      47060 ATACTTGATAATTCACCTGAGATAGGCCTTGTATATACTGATTATACTTTAATCGATGAG 47119

Qy        394 TTAAACTATAATGCTG 409
              |    |||  | ||
Db      47120 CAAGGTAATATAGGTG 47135


RESULT 6
AE017308_5
WPCOMMENT
  Sequence split into 8 fragments  LOCUS AE017308 Accession AE017308
     Fragment Name      Begin     End
     AE017308_0             1    110000
     AE017308_1        100001    210000
     AE017308_2        200001    310000
     AE017308_3        300001    410000
     AE017308_4        400001    510000
     AE017308_5        500001    610000
     AE017308_6        600001    710000
     AE017308_7        700001    777079
Continuation (6 of 8) of AE017308 from base 500001 (AE017308 Mycoplasma mobile 163K comp

  Query Match             6.5%;  Score 62.8;  DB 1;  Length 110000;
  Best Local Similarity   50.9%;  Pred. No. 0.19;
  Matches  176;  Conservative    0;  Mismatches  167;  Indels    3;  Gaps    1;

Qy         20 TTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTGGCTCGTTTACTTGATAGTA 79
              | | | ||  ||||||    |||||  ||      || | | |        | |  ||  ||
Db      11509 TAAGTGTTATTATTCCTTGTTATAATTGCGAAAAATTTATTGAAAAAGCAATAAATTCTA 11568

Qy         80 TAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTTATTGTTTGTGATAATGCTT 139
              |  || || |  | ||  |      |||  |  | ||||| || ||    || |||  |
Db      11569 TTTTTGAAGAAAATAAAAAAAAACATTGACAATATGGAGGTTCTTATTGTAAATGATGGAT 11628

Qy        140 CAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGATAAAATAAGAAATAGTACTT 199
              ||| |||| |  | | ||  |||   ||         |||||||     ||||  ||
Db      11629 CAAAAGATAATTCCGAAACTATACTTAACGAGTTGAAAGATAAATATCTAAATCTTAAAA 11688

Qy        200 ATCATCTAAATGAAGA---AAACTTAGGAATGGATGGTAACTTCCAGAAATGTTTTGAGT 256
              | ||  ||| ||| |    ||| | |||||  | |  ||| ||    ||| |  | | |
Db      11689 TTTATAAAAAAGAAAATGGAAATTGAGGAAGTGTTATTAATTTTGTAAAAAATAATAAAT 11748

Qy        257 TATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTAATAGTCAAAAATGGTA 316
              || |||  |||||||||  | |  ||  | |  | |||||||  |    | ||   ||
Db      11749 TAACAACAGGAAAATATTTATCTATTTTAGATGCTGATGATTGATATATAAAGGGTGCAT 11808

Qy        317 TTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGATAT 362
              |    ||  | |   | || ||  |          ||  |||  |||
Db      11809 TAGAAAAAATCTGACCTATTTTTGAAAAAAATTATGATTTAATTAT 11854


RESULT 7
AY253301
LOCUS       AY253301               15000 bp    DNA     linear   BCT 02-APR-2003
DEFINITION  Actinobacillus suis serotype K1 capsular gene cluster, complete
            sequence.
ACCESSION   AY253301
VERSION     AY253301.1  GI:29469153
KEYWORDS    .
SOURCE      Actinobacillus suis
  ORGANISM  Actinobacillus suis
            Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
            Pasteurellaceae; Actinobacillus.
REFERENCE   1  (bases 1 to 15000)
  AUTHORS   Slavic,D. and MacInnes,J.I.
  TITLE     Capsular gene cluster of Actinobacillus suis K1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 15000)
  AUTHORS   Slavic,D. and MacInnes,J.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2003) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .15000
                     /organism="Actinobacillus suis"
                     /mol_type="genomic DNA"
                     /strain="SO4"
                     /serotype="K1"
                     /db_xref="taxon:716"
     gene            complement(55. .705)
                     /gene="wzt"
     CDS             complement(55. .705)
                     /gene="wzt"
                     /codon_start=1
                     /transl_table=11
                     /product="Wzt"
                     /protein_id="AAO65487.1"
                     /db_xref="GI:29469154"
                     /translation="MISVKDVSKRYHTNSGWKTVLQNINFDLHKGEKIGILGRNGAGK
                     STLIRLMSGVEPPTTGTIERHMSISWPLAFSGAFQGSLTGMDNLRFICRIYNADFDYV
                     KAFTEEFSELGDYLYEPVKKYSSGMKARLAFALSLSVEFDCYLIDEVIAVGDSRFAAK
                     CKHELFEKRKDRSIILVSHSPSAMKEYCDNAMVLDKGIMHKFENMDEAYKFYNATQ"
     gene            complement(702. .1499)
                     /gene="wzm"
     CDS             complement(702. .1499)
                     /gene="wzm"
                     /codon_start=1
                     /transl_table=11
                     /product="Wzm"
                     /protein_id="AAO65488.1"
                     /db_xref="GI:29469155"
                     /translation="MQYGDQTTFRQSLAIQGRVIYALLMREIITRYGRKNLGFLWLFI
                     EPLLLTFLIVLMWKFFRADKVSTLNIVAFTITGYPMAMMWRNVSNRAIGSISANLSLL
                     YHRNVRVLDTIFARMLLEVAGATVAQILITAVLVFIGWIDPPKDVFYMVLAWTLMAFF
                     AFGLGLIICSLAQKIEVFGKIWNTLSFVLLPLSGAFFFVHSLPSQIREIAQWVPMISG
                     TEMFRHGYFGDLVPTYENIGFLVVCDLAMLLLGLILVRNFSKGIEPQ"
     gene            complement(1499. .2638)
                     /gene="wzf"
     CDS             complement(1499. .2638)
                     /gene="wzf"
                     /codon_start=1
                     /transl_table=11
                     /product="Wzf"
                     /protein_id="AAO65489.1"
                     /db_xref="GI:29469156"
                     /translation="MVAITAEKPNKPKKKSLLRKLNPLLWLTVIFPTLFSAIYFGLFA
                     SDIYVSESSFVVRSPRNQSSLSGVGALLQGTGFTRSQDDTYSVQEYMRSRTALEQLQT
                     ELPVRDFYATKGDILSRFNGFGLNDTQEAFFRYFKDRLSIDVDSVSGIATLRVQAFEA
                     SEGQQLNEKLLRLGEDLINRLNARGRKDTLEFAAQAVQEAEKNVNETAEALSKYRIKN
                     KIFDLPAQSGVQLSLISSLKSELIRVETQLAQLQSITPDNPQVDALLMRQKSLRKEID
                     EQSKQLSGNTGSSVANQSADYQRLVLANELAQQQLTAAMASLHNTRGEADRQQLYLEV
                     ISQPSKPDWALAPSRLYNIIATFIIGLMLYGIFGLLITSVREHKN"
     gene            complement(2688. .3869)
                     /gene="wza"
     CDS             complement(2688. .3869)
                     /gene="wza"
                     /codon_start=1
                     /transl_table=11
                     /product="Wza"
                     /protein_id="AAO65490.1"
                     /db_xref="GI:29469157"
                     /translation="MKPIKFRLLLSGLVASLAACSSLPTSGPSHSAILEANSQNSDKP
                     LPEVNLVELDNGLVQQLYQTQQSQQFSGFLGTAGGAGYAGAVNVGDVLEISIWEAPPA
                     VLFGGTFSSEGQGSGHLTQLPAQMVNQNGTVTVPFVGNIRVAGKTPEAIQSQIVGALQ
                     RKANQPQVLVKIANNNSADVTVIRQGNSIRMPLTANNERVLDAVAAVGGTTENIEDVT
                     VKLTRGSQVKTLAFETLISDPAQNIMLRAGDVVSLLNTPYSFTGLGAVGNNQQMKFSS
                     KGITLAEAIGKMGGLIDTRSDPRGVFVFRHVPFAQLSLDQQAQWEAKGYGIGMDVPTV
                     YRVNLLEPQSMFLLQRFPMQDKDIVYVSNAPLSEFQKFLRMIFSVTSPITSTTNSIRS
                     Y"
```

```
CDS             4105. .5217
                /note="ORF 1"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAO65491.1"
                /db_xref="GI:29469158"
                /translation="MNKMNRKFSKLLKNPHIFFRDFLNKKYPIKNTELPFAESEEANL
                IEAHQKLDKIIQKNTLRQSDIDVVFTWVDGSDPLWQSKYHQYAPNYQANSALYATDIA
                RFEDHNELYYSVHAVLKFMPWVRHIFIVTDNQKPKWLDEAYQEKITLISHQDIIDKEY
                LPTFNSHVIEAFLHKIPNLSENFIYFNDDVFIARELQAEHFFQANGIASIFMSEKSLT
                QMRDRGTITPTLSASEYSIRLLNKYYNTNIDSPLVHTYIPLKKSMYELAWQRYEKEIL
                GFLPNKLRTNHDLNFANFLIPWLMYFEGKAMPKIDICYYFNIRSPNALTQYKKLLNKK
                NIGEQPNSFCANDFNSQKSINNYQNQLFSFLNSYYS"
CDS             5228. .8554
                /note="ORF 2"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAO65492.1"
                /db_xref="GI:29469159"
                /translation="MNKVKRKFRKLLRDPKLFFSDMYFKHSIKIKKHLPVKYEGKHQF
                TIVSAVYNVEKYLDDFFDSIVKQNLSFKKHIQIILVDDGSKDSSANIIKKWQKKYPNN
                IHYYYKENGGQASARNLGLKYVQTEWVTFIDPDDFISNNYFRELDNFLSMNNELVLVG
                IPLIFYFEDKKIYKDTHPLKYRFSNGNKIFPTNDLENFIQLSASTALFKTRLINNIKF
                DEEMKPSFEDAKFVMDYLITNRIIGKVGFISDIQYFYRKREDGSSTLDGAWQNKNLFS
                KVLKHGCLSILESAEQQFGKIPTYIQRSVLYHLYWYFGRIVNNENALSHLTEYEKLEF
                SQIVHRIFNKIDKSTIEKFNLGGAWFYHKVGFLGLFKKAEPSFQIAYIKKFDLIKNQV
                LISFFSYQDANYSLLINGEDRMPSFKKDIHYNFLNERFTNEYRIWLNVQELGNLSILL
                NGKIAKLSFNGKWYNTLSMKVVREFYQSKSTKKENSWIFIDRDNQADDNAEHLYRYVM
                NNHPEKDIYFALNKNSKDWDRLNKEGFKLLEFKSKEFENKLKNCSKIISSHIDGYITH
                YFGDNSLLDKDYIFLQHGVTKDDMSPWLNTKENISIFITTTKDEYNSISEDGSPYRFT
                KKEVKFLGFPPRYDSLLSKNTFNTKNILVMPTWRQNIIGNSVNGSKRNFNSDFMETNYA
                KHWHNFLNGDMVKKLVNQYGYNIIFAPHPNIQEYLDVFTIPKYINTWRYSEGNIQKLF
                QEGMMLITDYSSVAFDMAYLEKYTIYYQFDEKEVFSGSHTYRKGYFEYDKHGFGPVAR
                TEDELNAILERFLINKSDNFDYIYSSRIRNTFIHRDQDNCKRVYEAIVKLDSNIPDEQ
                KFCYEIILESIKKAYQNEAWGVIYNRVNYLSKLNLIREEDKEWVTRIYLASIIKIRHF
                NEIEKFIPNNKYSDIILSLYYETHQHEKALELLSQSPEPDLFMLLYSYSNLYDFFNAK
                KISIKMLNKVNENLYPIILAYVSASQKNWENVIFLLSKKISMFTKEELKKYEPQLLLA
                KSYRHLKRYNEAHNMLVAFEKHTKDCSRCRIEISHLAYERADYKKCIDQLNKVFKFSL
                EYLPEESKRKYIESKNKLQK"
CDS             8595. .9827
                /note="ORF 3"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAO65493.1"
                /db_xref="GI:29469160"
                /translation="MKEMNFNATNNIVVKYKHKKSKYDFNHVIFVFSGFLNASPGNYD
                FSNALNDCPCDIIWINDEFEKMYTYYMCINMDFKVEEAITEFIYSKISELGLNKNQAT
                LTGFSKGGSAALYYGLKLNFSNIVVSVPQMKIGSYIENNWKQVASHMMGKNYTIVDKN
                YLDNILYKLLCQDTFLSRNIYLLTSEKDVQYSTEIVPYLSFFQKYTNFNLLKTHSAFV
                REHNQVTSHHVPLLLSIYYALATDAIPTYSGGEVNFFGRLLFSDKNPTNEMVIDLRVA
                KIINSHLFLEGVSFLQGNDLIEYSDVNYYLVLKLGESNIKLDLAKAHRPALTREFFNG
                KSLTIYDKAWFTTYQYKGIDISVLPKGKYQLSLGIQLSKGLSKVSVLKDSRNIVRTDT
                ENKYKLLSENNILYLEIL"
CDS             9841. .11490
                /note="ORF 4"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAO65494.1"
                /db_xref="GI:29469161"
                /translation="MTHYQIENCLENWDFDAALEWHRYISDNNSDQIPNYCRYLADTG
                HLDIAKSLIDSIKSNPVLYKKLSNDENFSVCRNLNSFFNKKLNEYANNPDYFCKLYMF
                SLTGNIDKVFSLLTTYRGGHINSSTSAENNMIINFALNKLIEKNRLDVDISREIIIHL
                ANSNKINNQRKKYLLKSMIDFIAKNHDLSKELFDLKHIYTIHIRLIPLIYAISNNENG
                AKSLMSKVYSLIQDNNNLNMLNTEKPRIAICISGMFKSDLTNLKTIQTKLAIPLNADV
                FIHTWDRQQDWMGDVRRYNFWPRVFNISNSLVPKNIQNLSFLEKNYSNVYSCLLSSVF
                SSLDINQVKNNIISKSILIENENNFMREHHINDNFKSRETFNQIKMFYGLYKCFELAK
                RKEDIEGFRYDYFIRLRADTIVNSNSISPEHLYALDNSSLAVPAGAGWGISDGFFYAN
                RSVYERVISLWKKMKIANRLSPFEEFRDWDAHKLLGLWLLKNDIRPVPCKFSCGTIFG
                GETLKVPGLLAALEKDNTQENRNKFPEETKWLMEFLKDKAK"

CDS             11552. .12268
                /note="ORF 5"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAO65495.1"
                /db_xref="GI:29469162"
                /translation="MFIIPMAGLSSRFFKAGYTKPKYQLEIGNETVFSWSVRSFERYF
                TTDKFIFIYRDVYETQDFLKQEIEKLGISDYELICLPEETLGQADTVYQGINHLPSDE
                EIYIFNIDSKIIEFIKPEWVNECDGYLEVFKGEGDHWSFALAENNSKRVIRTTEKERI
                SELCSDGLYYFKKKSIFESLFLNARKNNITSKNEYYIAPLYNDLIKQNGLVLYDLIEK
                EDILFCGTPDEYITLLGNEK"
CDS             12265. .12651
                /note="ORF 6"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAO65496.1"
                /db_xref="GI:29469163"
                /translation="MKRLIMDLDNTITLTENGDYRNAKPILDVIEKLREYKKQGFEII
                ISSSRNMRTYEGNVGKINVNTLPIIIDWLNRHNVPYDEIYVGKPWCGHDGFYVDDRAI
                RPDEFATMSYEEIRKLTKMDGEYDSN"
CDS             12629. .14197
                /note="ORF 7"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAO65497.1"
                /db_xref="GI:29469164"
                /translation="MVNMILINSAAYVNAEFRNEFGAIPPCFLPIGNRKLLTFQVNAL
                RENLGNDQHIVVSLPMNYSLSIDEKELIQNLNIQPVFIPEGISLGMAVLYVLNTVGHN
                EEVLRLLHGDTLLNNIPLENDCIALVSTQEDYEWEVDSNTNIPLVWCGYFSFTSSQKF
                IKALATTQGDFVQSVHMYAKEEPNTIHKEVNEWYDLGHINTYFRSRSSITTQRAFNSL
                KIGNGVVWKSGTPARKIEAEAHWFANLPAQLKRFTPQLINAGVLKGNPFYETEYLPIL
                PLNEIFVHGKNPVAFWEKVLNLITFYMSESRAHLPKENKDLLEKIHVDSLSLYSDKTY
                ERLEKYAEQSGIDLDQSTRYNGIDLPSLREIATECVTRTLKLPEIPAIVHGDLCFSNV
                MYDSRSNNIKVIDPRGLNIQQELTIYGNQSYDLAKLCHSFIGLYDFIIADSFKLEKSE
                SLGVKLHFNIDSRLEAIQDVFMRKTLLPNISNKDIIAPTILLFLSMIPLHFDKPHRQE
                AMLANALRLYSEWL"
ORIGIN

  Query Match             6.4%;  Score 61.6;  DB 1;  Length 15000;
  Best Local Similarity   49.1%;  Pred. No. 0.65;
  Matches  163;  Conservative    0;  Mismatches  169;  Indels    0;  Gaps    0;

Qy          9 AAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTGGCTCGTTT 68
              ||||| ||   |||| |||| |    |   ||||||          || | | | |  |||
Db       5347 AAAACATCAATTTACGATTGTTTCCGCTGTATATAATGTAGAAAAATATCTTGATGATTT 5406

Qy         69 ACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTTATTGTTTG 128
                ||||||||||  || ||||  |     |  | | |  |  |  |  || ||||  |
Db       5407 CTTTGATAGTATCGTTAAACAAAATTTATCATTTAAAAAACACATACAGATTATCTTAGT 5466

Qy        129 TGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGATAAAATAAG 188
              |||| | |  |||| ||||  | ||||||  ||   ||| |    |   | | |||
Db       5467 TGATGACGGCTCAAAAGATTCATCAGCAAACATCATCAAAAAATGGCAAAAAAAATATCC 5526

Qy        189 AAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTCCAGAAATG 248
              ||||| ||   |  ||    || |||||||    ||    |   |       |   | |
Db       5527 AAATAATATCCACTATTATTATAAAGAAAATGGTGGGCAAGCCTCTGCTCGTAATTTAGG 5586

Qy        249 TTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTAATAGTCAA 308
                | | |    | |   | |||   ||     | ||||      ||||||  | ||   |||
Db       5587 ACTAAAATACGTACAAACAGAATGGGTTACCTTTATTGATCCGGATGATTTTATTAGCAA 5646

Qy        309 AAATGGTATTTCGAAGGTTTTTTCGATATTAA 340
               ||     ||    || |   |    | ||||
Db       5647 CAACTACTTTAGAGAGCTAGATAATTTCTTAA 5678


RESULT 8
AF461769
LOCUS       AF461769               19457 bp    DNA     linear   BCT 26-SEP-2002
```

```
DEFINITION  Yersinia pseudotuberculosis serogroup IVB O-antigen gene cluster,
            complete sequence.
ACCESSION   AF461769 U29691
VERSION     AF461769.1  GI:23321111
KEYWORDS    .
SOURCE      Yersinia pseudotuberculosis
  ORGANISM  Yersinia pseudotuberculosis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Yersinia.
REFERENCE   1  (bases 1 to 19457)
  AUTHORS   Hobbs,M. and Reeves,P.R.
  TITLE     Genetic organisation and evolution of Yersinia pseudotuberculosis
            3,6-dideoxyhexose biosynthetic genes
  JOURNAL   Biochim. Biophys. Acta 1245 (3), 273-277 (1995)
   PUBMED   8541300
REFERENCE   2  (bases 1 to 19457)
  AUTHORS   Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C.,
            Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
  TITLE     Bacterial polysaccharide synthesis and gene nomenclature
  JOURNAL   Trends Microbiol. 4 (12), 495-503 (1996)
   PUBMED   9004408
REFERENCE   3  (bases 1 to 19457)
  AUTHORS   Pacinelli,E., Wang,L. and Reeves,P.R.
  TITLE     Relationship of Yersinia pseudotuberculosis O antigens IA, IIA, and
            IVB: the IIA gene cluster was derived from that of IVB
  JOURNAL   Infect. Immun. 70 (6), 3271-3276 (2002)
   PUBMED   12011023
REFERENCE   4  (bases 1 to 19457)
  AUTHORS   Hobbs,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-1995) Department of Microbiology, University of
            Sydney, Building G08, Sydney, NSW 2006, Australia
REFERENCE   5  (bases 1 to 19457)
  AUTHORS   Pacinelli,E., Wang,L. and Reeves,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) Microbiology, University of Sydney, Sydney,
            NSW 2006, Australia
  REMARK    Nucleotide and amino acid sequences updated by submitter
COMMENT     On Sep 26, 2002 this sequence version replaced gi:1223958.
FEATURES             Location/Qualifiers
     source          1. .19457
                     /organism="Yersinia pseudotuberculosis"
                     /mol_type="genomic DNA"
                     /isolate="M452"
                     /db_xref="taxon:633"
                     /note="serogroup: IVB"
     gene            243. .1232
                     /gene="ddhD"
     CDS             243. .1232
                     /gene="ddhD"
                     /note="DdhD"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-6-deoxy-delta-3,4-glucoseen reductase"
                     /protein_id="AAN23052.1"
                     /db_xref="GI:23321112"
                     /translation="MSLNVKLHPSGIIFTSDGTSTILDAALDSNIHIEYSCKDGTCGS
                     CKAILISGEVDSAENTFLTEEDVAKGAILTCCSKAKSDIELDVNYYPELSHIQKKTYP
                     CKLDSIEFVGEDIAILSLRLPPTAKIQYLAGQYIDLIINGQRRSYSIANAPGGNGNIE
                     LHVRKVVNGVFSNIIFNELKLQQLLRIEGPQGTFFVREDNLPIVFLAGGTGFAPVKSM
                     VEALINKNDQRQVHIYWGMPAGHNFYSDIANEWAIKHPNIHYVPVVSGDDSTWTGATG
                     FVHQAVLEDIPDLSLFNVYACGSLAMITAARNDFINHGLAENKFFSDAFVPSK"
     gene            1258. .2043
                     /gene="ddhA"
     CDS             1258. .2043
                     /gene="ddhA"
                     /note="DdhA"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose-1-P-cytidylyltransferase"
                     /protein_id="AAN23053.1"
                     /db_xref="GI:23321113"
                     /translation="MEIQVKAVILAGGLGTRLSEETVVKPKPMVEIGGKPILWHIMKL
                     YSSYGINDFVICCGYKGYVIKEYFANYFMHMSDITFCMRDNEMIVHQKRVEPWNVTLV
                     DTGEDSMTGGRLRRVKDYVKDDEAFCFTYGDGVSDVNIAELIEFHKSHGKQATLTATY
                     PPGRFGALDIKDKQVRSFKEKPKGDGALINGGYFVLSPKVIDLIDGDKSTWEQEPLMT
                     LAAQGELMAFEHAGFWQPMDTLRDKIYLHELWEEGRAPWKVWE"
     gene            1946. .3121
                     /gene="ddhB"
     CDS             1946. .3121
                     /gene="ddhB"
                     /note="DdhB"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-glucose-4,6-dehydratase"
                     /protein_id="AAN23054.1"
                     /db_xref="GI:23321114"
                     /translation="MPVSGSQWIRCVTRSICMNYGKKAGHLGRYGNNKMINNSFWQGK
                     RVFVTGHTGFKGGWLSLWLQTMGATVKGYSLAPPTVPSLFETARVADGMQSEIGDIRD
                     QNKLLESIREFQPEIVFHMAAQPLVRLSYSEPVETYSTNVMGTVYLLEAIRHVGGVKA
                     VVNITSDKCYDNKEWIWGYRENEAMGGYDPYSNSKGCAELVTSSYRNSFFNPANYGQH
                     GTAVATVRAGNVIGGGDWALDRIVPDILRAFEQSQPVIIRNPHAIRPWQHVLEPLSGY
                     LLLAQKLYTDGAEYAEGWNFGPNDADATPVKNIVEQMVKYWGEGASWQLDGNAHPHEA
                     HYLKLDCSKAKMQLGWHPRWNLNTTLEYIVGWHKNWLSGTDMHEYSITEINNYMNTK"
     gene            3139. .4452
                     /gene="ddhC"
     CDS             3139. .4452
                     /gene="ddhC"
                     /note="DdhC; formerly known as rfbH"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-4-keto-6-deoxy-D-glucose-3-dehydrase"
                     /protein_id="AAN23066.1"
                     /db_xref="GI:23321126"
                     /translation="MSQEELRQQIAELVAQYAETAMAPKPFEAGKSVVPPSGKVIGTK
                     ELQLMVEASLDGWLTTGRFNDAFEKKLGEYLGVPYVLTTTSGSSANLLALTALTSPKL
                     GVRALKPGDEVITVAAGFPTTVNPTIQNGLIPVFVDVDIPTYNVNASLIEAAVSDKTK
                     AIMIAHTLGNLFDLAEVRRVADKYNLWLIEDCCDALGSTYDGKMAGTFGDIGTVSFYP
                     AHHITMGEGGAVFTQSAELKSIIESFRDWGRDCYCAPGCDNTCKKRFGQQLGSLPFGY
                     DHKYTYSHLGYNLKITDMQAACGLAQLERIEEFVEKRKANFKYLKDALQSCADFIELP
                     EATENSDPSWFGFPITLKEDSGVSRIDLVKFLDEAKVGTRLLFAGNLTRQPYFHDVKY
                     RVVGELTNTDRIMNQTFWIGIYPGLTHDHLDYVVSKFEEFFGLNF"
     repeat_region   4463. .4478
                     /rpt_type=inverted
     gene            4488. .5345
                     /gene="prt"
     CDS             4488. .5345
                     /gene="prt"
                     /note="Prt; formerly known as rfbS"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-paratose synthetase"
                     /protein_id="AAN23067.1"
                     /db_xref="GI:23321127"
                     /translation="MKILITGVSGYLGSQLANALMLEHEVAGTVRAGSVCNRITDIGN
                     VNLINVTDSGWIDKVLSFSPDVVINTVALYGRKGELLSELVDANIQFPLRILEMLVST
                     GKGLFFQCGTSLPADVSQYALTKNQFTELAREYCNKFSGKFIELKLEHFFGPFDDSTK
                     FTTYVINSCRSHSDLKLTAGLQRRDFIYINDLINAFKIMISKSESLISGESISIGSGH
                     AVTIKEFVETVAKMTSYQGNLQFGAIPTRENELMYSCASLARIQELGWLCQYSLNSAI
                     KDILHKEYK"
     gene            5342. .6358
                     /gene="tyv"
     CDS             5342. .6358
                     /gene="tyv"
                     /note="Tyv; formerly known as rfbE"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-tyvelose epimerase"
                     /protein_id="AAN23068.1"
                     /db_xref="GI:23321128"
                     /translation="MKLLITGGCGFLGSNLASHAIKSGMEVIVFDNLSRYGSSDNLKW
                     LQSIGGFTYVHGDIRNRNDITRLIQKFKPDNIFHLAGQVAMTTSIDNPQMDFEVNVGG
                     TFNLLEAIRLFNPECGIIYSSTNKVYGDLEQFTYRETDTRYECIEMPDGFDESTQLTF
                     HSPYGCSKGAADQYMLDYARIYGLKTVVFRHSSMYGGRQFSTYDQGWVGWFCQKAIEA
                     SRGVNSPFTISGNGKQVRDVLHAEDIISLYFSTLSNLERVKGNAFNIGGTIEHSLSLL
                     ELFSLLEKYTETELKYTRIPVRESDQKVFVANINKISESTGWIPKVSSESGIKIMLDW
```

```
                VETV"
     gene       6390. .7703
                /gene="wzx"
     CDS        6390. .7703
                /gene="wzx"
                /note="Wzx; formerly known as rfbX"
                /codon_start=1
                /transl_table=11
                /product="O-unit flippase-like protein"
                /protein_id="AAN23069.1"
                /db_xref="GI:23321129"
                /translation="MRVPTHIIVAASAWGSRLVSIFIQFYSIKILLDLLGTEGYAVFT
                VIGSLVGWFLLADFGLGNSLQNQISYRRANHQEYQDLVLSAVIAIIPIFILFIILILT
                LSPYISEFLLGGFDFLNNNQRSNIFKAASFIFLTTSIGNLAYKIWFSEHKGWISNIIP
                ALSSIVGLIFLMRLPSDGSNISEDIIFSIYCFYIPAAFFGVTSTLFKVIPYLKCKNFL
                NKLTLYALIKNGGGFFLFSLLSAFVLQVDYIVMSQTLVERDLVTYNIMSKTFGLINFI
                YAALLQSLWPVCAEASSKLRFDNFYKIEKKYISFGFIIVIVSSFVIFLLKDFIVNILA
                PGKDFYFPISLIILFSFYQVVRVWTDTYAMFLMSIGKLKPLWISVPFQAVLSGSLQWV
                GAVNYGLVGLLCGLIASFLITVSWWLPFSFRSTVDRIVKDKRLDE"
     gene       7696. .8733
                /gene="wbyA"
     CDS        7696. .8733
                /gene="wbyA"
                /note="WbyA"
                /codon_start=1
                /transl_table=11
                /product="putative tyvelose transferase"
                /protein_id="AAN23055.1"
                /db_xref="GI:23321115"
                /translation="MSNIKISFCIPTYNRSELLSELIESIVSQCKNRNDIEICISDNA
                SSDNTSEMIAKWKESTDTPIVYKVNKENIGPDRNYLAAVDLASGEYCWLFGSDDRLLS
                DAIIKMDEHINEGHDIYIVDRIECDFNMKPFSHRTWMSTGSKLYNTSKRTDLIEYFNN
                SYAVGAVFSYLSSIIVKRSRWNRVKFNPKFIGSCYAHVFYLLDIVNNGAELKYVKSAF
                VLSRGGNDSFAANGLIKRIEIDFDGYFLLGNEFFSNDFYLNSKFKAILLKEHPVIHTC
                IKTAMIGEFQDLNKIKRFYIELGIAPYYVSIIYFFKPIYQLANKMKLSFFYKKLRKKI
                TFKKGSFNSID"
     gene       8768. .9877
                /gene="wbyB"
     CDS        8768. .9877
                /gene="wbyB"
                /note="WbyB"
                /codon_start=1
                /transl_table=11
                /product="putative 6-deoxy-D-mannoheptose transferase"
                /protein_id="AAN23056.1"
                /db_xref="GI:23321116"
                /translation="MNILIVYLGRKGGGAHYTYELVNELSKKTNVSLLVSKQAENIKS
                YESLNISLNTINTFNGMFSFILSSFKLPFIKNKIKKIILDEKIDIVISTMTHVWNGFI
                FTPNFMMKIPYILTIHDAVPHSGEENFILNLLNNKDIKNANAYIALTEHVKNQFINIY
                GDKKDICILPLGSFNFKSGEIKSWCGVRPLRILFFGRIHEYKGIDLLLDSMIILQEKY
                DLELSIYGQGDLSDYVDKIDVIKKINIENRWIGDSEVGDILQAHDVCVIPYKDASQSG
                VIPVAQSAGLPTIVNPVDGLIEQIEKDKTSLVTDRVDSVSLALSIEKIIKNPELVNKL
                SIGSLEYASRKLSWVPIVNKLITYSKETLNSSDRS"

  Query Match             6.3%;  Score 60.8;  DB 1;  Length 19457;
  Best Local Similarity   48.5%;  Pred. No. 0.79;
  Matches  167;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy         26 TTGCTATTCCGACATATAACCGCTCTTCATGTTTGGCTCGTTTACTTGATAGTATAATTC 85
              ||  |||||| || ||||| || |||    |  ||  |      || |||| |||||  ||
Db       7718 TTTGTATTCCTACGTATAATCGTTCTGAGTTATTATCAGAACTAATTGAAAGTATTGTTT 7777

Qy         86 AACAGGAGAACTATTGTCATGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAG 145
                |||    ||  || |  |||||        ||||     | ||    |||||
Db       7778 CACAATGTAAAAATAGAAATGATATTGAAATATGTATTTCAGATAATGCATCTTCAGATA 7837

Qy        146 ATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATC 205
              ||  | | | ||  || || ||  |          ||      |      |    |||| |
Db       7838 ATACATCTGAAATGATTGCTAAATGGAAAGAGAGTACTGACACTCCAATAGTTTATAAAG 7897

Qy        206 TAAATGAAGAAAACTTAGGAATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATG 265
              ||||| | |||||| | ||    ||| | || |  |          | || ||  |   ||
Db       7898 TAAATAAGGAAAACATTGGGCCAGATAGAAATTATCTTGCTGCCGTAGACTTGGCTTCTG 7957

Qy        266 GAAAATATCTTTGGATGATTGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGG 325
              |  |||||  |||| || ||||   |||||||  | |  |      |||  ||    ||
Db       7958 GTGAATATTGTTGGTTGTTTGGTAGTGATGATAGATTGCTATCTGATGCAATAATTAAAA 8017

Qy        326 TTTTTTCGATATTAAAGTCCCGGCCTGCATTAGATATGGTGTAT 369
              |   |        | ||     | | ||   |  |||| ||  ||
Db       8018 TGGATGAACATATCAATGAAGGACATGATATCTATATCGTAGAT 8061


RESULT 9
AF461770
LOCUS       AF461770               20125 bp    DNA     linear   BCT 26-SEP-2002
DEFINITION  Yersinia pseudotuberculosis serogroup IIA O-antigen gene cluster,
            complete sequence.
ACCESSION   AF461770 L01777
VERSION     AF461770.1  GI:23321130
KEYWORDS    .
SOURCE      Yersinia pseudotuberculosis
  ORGANISM  Yersinia pseudotuberculosis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Yersinia.
REFERENCE   1  (bases 1 to 20125)
  AUTHORS   Kessler,A.
  TITLE     The 3,6 dideoxyhexose pathway genes of Yersinia pseudotuberculosis
            serogroup IIA
  JOURNAL   Thesis (1993) Department of Microbiology, University of Sydney,
            Australia
REFERENCE   2  (bases 1 to 8845)
  AUTHORS   Kessler,A.C., Haase,A. and Reeves,P.R.
  TITLE     Molecular analysis of the 3,6-dideoxyhexose pathway genes of
            Yersinia pseudotuberculosis serogroup IIA
  JOURNAL   J. Bacteriol. 175 (5), 1412-1422 (1993)
   PUBMED   8444803
REFERENCE   3  (bases 1 to 20125)
  AUTHORS   Hobbs,M. and Reeves,P.R.
  TITLE     The JUMPstart sequence: a 39 bp element common to several
            polysaccharide gene clusters
  JOURNAL   Mol. Microbiol. 12 (5), 855-856 (1994)
   PUBMED   8052136
REFERENCE   4  (bases 1 to 20125)
  AUTHORS   Liu,D., Lindqvist,L. and Reeves,P.R.
  TITLE     Transferases of O-antigen biosynthesis in Salmonella enterica:
            dideoxyhexosyltransferases of groups B and C2 and acetyltransferase
            of group C2
  JOURNAL   J. Bacteriol. 177 (14), 4084-4088 (1995)
   PUBMED   7541787
REFERENCE   5  (bases 1 to 20125)
  AUTHORS   Hobbs,M. and Reeves,P.R.
  TITLE     Genetic organisation and evolution of Yersinia pseudotuberculosis
            3,6-dideoxyhexose biosynthetic genes
  JOURNAL   Biochim. Biophys. Acta 1245 (3), 273-277 (1995)
   PUBMED   8541300
REFERENCE   6  (bases 1 to 20125)
  AUTHORS   Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C.,
            Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
  TITLE     Bacterial polysaccharide synthesis and gene nomenclature
  JOURNAL   Trends Microbiol. 4 (12), 495-503 (1996)
   PUBMED   9004408
REFERENCE   7  (bases 1 to 20125)
  AUTHORS   Pacinelli,E., Wang,L. and Reeves,P.R.
  TITLE     Relationship of Yersinia pseudotuberculosis O antigens IA, IIA, and
            IVB: the IIA gene cluster was derived from that of IVB
  JOURNAL   Infect. Immun. 70 (6), 3271-3276 (2002)
   PUBMED   12011023
REFERENCE   8  (bases 1 to 20125)
  AUTHORS   Hobbs,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1993) Department of Microbiology, University of
            Sydney, Building G08, Sydney, NSW 2006, Australia
REFERENCE   9  (bases 1 to 20125)
  AUTHORS   Pacinelli,E., Wang,L. and Reeves,P.R.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (21-DEC-2001) Microbiology, University of Sydney, Sydney,
            NSW 2006, Australia
  REMARK    Nucleotide and amino acid sequences updated by submitter
COMMENT     On Sep 26, 2002 this sequence version replaced gi:896325.
FEATURES             Location/Qualifiers
     source          1. .20125
                     /organism="Yersinia pseudotuberculosis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:633"
                     /note="serogroup: IIA"
     repeat_region   1. .10009
                     /note="remnants of an IS related to IS630"
                     /insertion_seq="IS630-like"
     gene            <1. .235
                     /gene="hemH"
     CDS             <1. .235
                     /gene="hemH"
                     /codon_start=2
                     /transl_table=11
                     /product="ferrochelatase"
                     /protein_id="AAB49397.1"
                     /db_xref="GI:155494"
                     /translation="TPYTDETLKSLPSQGVKHIQLICPGFSADCLETLEEIKEQNREF
                     FLHAGGEKFEYIPALNDDEGHIALLEQLIRHNI"
     misc_feature    586. .623
                     /note="JUMPstart"
                     /citation=[3]
     gene            833. .1822
                     /gene="ddhD"
     CDS             833. .1822
                     /gene="ddhD"
                     /note="DdhD; formerly known as rfbI"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-6-deoxy-delta-3,4-glucoseen reductase"
                     /protein_id="AAB49398.1"
                     /db_xref="GI:155495"
                     /translation="MSLNVKLHPSGIIFTSDGTSTILDAALDSNIHIEYSCKDGTCGS
                     CKAILISGEVDSAENTFLTEEDVAKGAILTCCSKAKSDIELDVNYYPELSHIQKKTYP
                     CKLDSIEFIGEDIAILSLRLPPTAKIQYLAGQYIDLIINGQRRSYSIANAPGGNGNIE
                     LHVRKVVNGVFSNIIFNELKLQQLLRIEGPQGTFFVREDNLPIVFLAGGTGFAPVKSM
                     VEALINKNDQRQVHIYWGMPAGHNFYSDIANEWAIKHPNIHYVPVVSGDDSTWTGATG
                     FVHQAVLEDIPDLSLFNVYACGSLAMITAARNDFINHGLAENKFFSDAFVPSK"
     repeat_region   1822. .1839
                     /rpt_type=inverted
     gene            1848. .2633
                     /gene="ddhA"
     CDS             1848. .2633
                     /gene="ddhA"
                     /note="DdhA; formerly known as rfbF"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose-1-P-cytidylyltransferase"
                     /protein_id="AAB49399.1"
                     /db_xref="GI:155496"
                     /translation="MEIQVKAVILAGGLGTRLSEETVVKPKPMVEIGGKPILWHIMKL
                     YSSYGINDFVICCGYKGYVIKEYFANYFMHMSDITFCMRDNEMIVHQKRVEPWNVTLV
                     DTGEDSMTGGRLRRVKDYVKDDEAFCFTYGDGVSDVNIAELIEFHKSHGKQATLTATY
                     PPGRFGALDIKDKQVRSFKEKPKGDGALINGGYFVLSPKVIDLIDGDKSTWEQEPLMT
                     LAAQGELMAFEHAGFWQPMDTLRDKIYLHELWEEGRAPWKVWE"
     gene            2536. .3711
                     /gene="ddhB"
     CDS             2536. .3711
                     /gene="ddhB"
                     /note="DdhB; formerly known as rfbG"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-glucose-4,6-dehydratase"
                     /protein_id="AAN23079.1"
                     /db_xref="GI:23321140"
                     /translation="MPVSGSQWIRCVTRFICMNYGKKAGHLGRYGNNKMINNSFWQGK
                     RVFVTGHTGFKGGWLSLWLQTMGATVKGYSLAPPTVPSLFETARVADGMQSEIGDIRD
                     QNKLLEAIREFQPEIVFHMAAQPLVRLSYSEPVETYSTNVMGTVYLLEAIRHVGGVKA
                     VVNITSDKCYDNKEWIWGYRENEAMGGYDPYSNSKGCAELVTSSYRNSFFNPANYGQH
                     GTAVATVRAGNVIGGGDWALDRIVPDILRAFEQSQPVIIRNPHAIRPWQHVLEPLSGY
                     LLLAQKLYTDGAEYAEGWNFGPNDADATPVKNIVEQMVKYWGEGASWQLDGNAHPHEA
                     HYLKLDCSKAKMQLGWHPRWTLNTTLEYIVGWHKNWLSGTDMHEYSITEINNYMNTK"
     gene            3729. .5042
                     /gene="ddhC"
     CDS             3729. .5042
                     /gene="ddhC"
                     /note="DdhC; formerly known as rfbH"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-4-keto-6-deoxy-D-glucose-3-dehydrase"
                     /protein_id="AAB49401.1"
                     /db_xref="GI:155498"
                     /translation="MSQEELRQQIAELVAQYAETAMAPKPFEAGKSVVPPSGKVIGTK
                     ELQLMVEASLDGWLTTGRFNDAFEKKLGEYLGVPYVLTTTSGSSANLLALTALTSPKL
                     GVRALKPGDEVITVAAGFPTTVNPTIQNGLIPVFVDVDIPTYNVNASLIEAAVSDKTK
                     AIMIAHTLGNLFDLAEVRRVADKYNLWLIEDCCDALGSTYDGKMAGTFGDIGTVSFYP
                     AHHITMGEGGAVFTQSAELKSIIESFRDWGRDCYCAPGCDNTCKKRFGQQLGSLPFGY
                     DHKYTYSHLGYNLKITDMQAACGLAQLERIEEFVEKRKANFKYLKDALQSCADFLELP
                     EATENSDPSWFGFPITLKEDSGVSRIDLVKFLDEAKVGTRLLFAGNLTRQPYFHDVKY
                     RVVGELTNTDRIMNQTFWIGIYPGLTHDHLDYVVSKFEEFFGLNF"
     repeat_region   5056. .5071
                     /rpt_type=inverted
     gene            5082. .5933
                     /gene="abe"
     CDS             5082. .5933
                     /gene="abe"
                     /note="Abe; formerly known as rfbJ"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-abequose synthase"
                     /protein_id="AAB49402.1"
                     /db_xref="GI:896326"
                     /translation="MRIVLTGGSGYIGSSLTPVLIKKYGRVYNIGRNTISEVSINGSK
                     EYCEFTYESLFDSLVELSPDLVINLAAGYYNDSGAPDLNVIDGNLKIPFIILEYFKSC
                     NYGRFINIGSYWEFSCSGRGVKGVNPYGIIKSTVRRLLDYYSKYNVIYTNLILYGSYG
                     DNDHRGKIVDCIIDAVNSNETLKLSPGEQKLNLVYIDDIIEAILYIVSSDNGQYDNET
                     LSIYTPTEHTVKEIVCFINEIKDNNLSLGGGRYRNDEVMAPDYKYRNIFHAKDKLKEY
                     ITSKIKK"
     misc_feature    6443. .6754
                     /note="part of non-functional tyv gene which has
                     accumulated numerous mutations"
     CDS             6755. .7375
                     /note="non-functional gene; orf6.8; remaining coding
                     sequence of mutated non-functional tyv gene"
                     /pseudo
                     /codon_start=1
                     /transl_table=11
     gene            7407. .8720
                     /gene="wzx"
     CDS             7407. .8720
                     /gene="wzx"
                     /note="putative O antigen export protein; Wzx; previously
                     called rfbX and orf7.4; TTG initiation codon"
                     /codon_start=1
                     /transl_table=11
                     /product="O-unit flippase-like protein"
                     /protein_id="AAB49403.1"
                     /db_xref="GI:409590"
                     /translation="MRVPTHIIVAASAWGSRLVSIFIQFYSIKILLDLLGTEGYAVFT
                     VIGSLVGWFLLADFGLGNSLQNQISYRRANHQEYQDLVLSAVIAIIPIFILFIILILT
                     LSPYISEFLLGGFDFLNNNQRSNIFKVASFIFLTTSIGNLAYKIWFSEHKGWVSNIIP
                     ALSSIVGLVFLMRLPSDGSNISEDIIFSIYCFYIPAAFFGVISTLFKVIPYLKCKNFL
                     NKLTLYTLIKNGGGFFLFSVLSALVLQVDYIVMSQTLVERDLVTYNIMSKTFGLINFI
                     YAALLQSLWPVCAEASSKLRFDNFYKIEKKYISFGFIIVIASSFVIFLLKDFIVNILA

  Query Match             6.3%;  Score 60.8;  DB 1;  Length 20125;
  Best Local Similarity   48.5%;  Pred. No. 0.78;
  Matches  167;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy         26 TTGCTATTCCGACATATAACCGCTCTTCATGTTTGGCTCGTTTACTTGATAGTATAATTC 85
              ||  |||||| || ||||| || |||   |  ||  |     || |||| |||||  ||
```

```
Db       8735 TTTGTATTCCTACGTATAATCGTTCTGAGTTATTATCAGAACTAATTGAAAGTATTGTTT 8794

Qy         86 AACAGGAGAACTATTGTCATGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAG 145

Db       8795 CACAATGTAAAAATAGAAATGATATTGAAATATGTATTTCAGATAATGCATCTTCAGATA 8854

Qy        146 ATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATC 205

Db       8855 ATACATCTGAAATGATTGCTAAATGGAAAGAGAGTACTGACACTCCAATAGTTTATCAAG 8914

Qy        206 TAAATGAAGAAAACTTAGGAATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATG 265

Db       8915 TAAATAAGGAAAATATTGGGCCAGATAGAAATTATCTTGCTGCCGTAGACTTGGCTTCTG 8974

Qy        266 GAAAATATCTTTGGATGATTGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGG 325

Db       8975 GTGAATATTGTTGGTTGTTTGGTAGTGATGATAGATTGCTATCTGATGCAATAATTAAAA 9034

Qy        326 TTTTTTCGATATTAAAGTCCCGGCCTGCATTAGATATGGTGTAT 369

Db       9035 TGGATGAACATATCAATGAAGGACATGATATCTATATCGTAGAT 9078
```

```
RESULT 10
RPXX02
LOCUS       RPXX02                312430 bp    DNA     linear   BCT 15-APR-2005
DEFINITION  Rickettsia prowazekii strain Madrid E, complete genome; segment
            2/4.
ACCESSION   AJ235271 AJ235269
VERSION     AJ235271.1  GI:3868717
KEYWORDS    complete genome.
SOURCE      Rickettsia prowazekii
  ORGANISM  Rickettsia prowazekii
            Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
            Rickettsiaceae; Rickettsieae; Rickettsia; typhus group.
REFERENCE   1
  AUTHORS   Andersson,S.G., Zomorodipour,A., Andersson,J.O.,
            Sicheritz-Ponten,T., Alsmark,U.C., Podowski,R.M., Naslund,A.K.,
            Eriksson,A.S., Winkler,H.H. and Kurland,C.G.
  TITLE     The genome sequence of Rickettsia prowazekii and the origin of
            mitochondria
  JOURNAL   Nature 396 (6707), 133-140 (1998)
   PUBMED   9823893
REFERENCE   2  (bases 1 to 312430)
  AUTHORS   Andersson,S.G.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1998) S.G.E. Andersson,
            Siv.Andersson@molbio.uu.se, Dept. of Molecular Biology, University
            of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN
COMMENT     On Nov 13, 1998 this sequence version replaced gi:3860788.
FEATURES             Location/Qualifiers
     source          1. .312430
                     /organism="Rickettsia prowazekii"
                     /mol_type="genomic DNA"
                     /strain="Madrid E"
                     /db_xref="taxon:782"
     gene            76. .1830
                     /gene="RP226"
     CDS             76. .1830
                     /gene="RP226"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA14689.1"
                     /db_xref="GI:3860789"
                     /db_xref="InterPro:IPR002110"
                     /db_xref="UniProt/TrEMBL:Q9ZDU8"
                     /translation="MLNKLCDILFLINLLLVTVQGYASPPPLPPSLPIIAVDTTDKNI
                     STNSNISFFEKFKQFFSKQKKKNISSQHEQEQTKAIHQESQQIDSRELNENEQSEPFI
                     DFGSTILPSVASNYIDSKAEYENSTNLAASYNTQDIQVKQQEFDPSEASEPIDIGNTK
                     FTSATNHEMYKEAVSSNDKETNLTSNIITPNVPSPVISIPTAQDVNYVVPSQQSVQIY
                     KPTNLTSIRNPIPLNHHTDLNKVEKNLESTISNMTTIPTNMVSVPSIQDTIQTTLNIT
                     VPTAETHVSVQTSTVMHSNQHSAQPITPISINTPVETSSTVLRATESSVPINNSQEIF
                     VSESESTKKQDWYTPIMPVLVVDPNKSQSKPLALEQKNNNDQIINNQAESHSVSSSNV
                     TIQKQNDKVNNATSESTKEFVKNETQMLFLPDDDIVLGKLTEDATLEQMDMHGYIKLF
                     QKKEEWIANAEKRKLVESFIKYDDDINKNKDIYANLSYYSAVDNAFRAVEKNNLFELR
                     ALLDVYPILQAKNSTGETLLTSSIYNGNYYLAKFLVIRGIKTSVLKNDECKYPLDIAL
                     AKGNTNIVCMLMKAKGYN"
     gene            2025. .4013
                     /gene="RP227"
     CDS             2025. .4013
                     /gene="RP227"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA GYRASE SUBUNIT B (gyrB1)"
                     /protein_id="CAA14690.1"
                     /db_xref="GI:3860790"
                     /db_xref="GOA:Q9ZDU7"
                     /db_xref="InterPro:IPR001241"
                     /db_xref="InterPro:IPR002288"
                     /db_xref="InterPro:IPR003594"
                     /db_xref="InterPro:IPR005737"
                     /db_xref="InterPro:IPR011558"
                     /db_xref="UniProt/TrEMBL:Q9ZDU7"
                     /translation="MSDLFSLNKEKKNKIVYTNYSAKDIEVLDGLEPVRKRPGMYIGG
                     TDSNAMHHLVSEVLDNAMDEAVAGFASIIMIKMHQDHSITIFDNGRGIPIDNHPKFPD
                     KSALEVILTTLHSGGKFSNNVYHTSGGLHGVGISVVNALSKHFKIKVYKQGKLYSQSY
                     SKGAKLTDLISAEASKRLRGTSINFTPDPEIFSEKLHFNPKKIYEIARSKAYLYRGVS
                     IEWECEVEVPSDIPKKALINFPNGLKDYLSSKISLDNLVIPEIFSGNIESTVDDIKLE
                     WAICWQNNDTSAFMQSYCNTVPTPQGGTHEQGLKSAILRGLKAYSEMIGNKKAANLTI
                     EDILETASIVLSIFIVEPSFQGQTKEKLVSNGVSKLVENIIKDHFDHFLSSDKVLATH
                     LLEHVIAIAEFRRNKKNERNISRKSVTQKLRLPGKLADCTRTSAEGTELFIVEGDSAG
                     GSAKQARNRETQAVLPLWGKVLNVASSTLEKIINNQAIQDLEIALACGSLKNYKKENL
                     RYEKIIIMTDADVDGAHIASLLMTFFFLRMPKLVEEGHLYLAKPPLYRLTQSNKIYYA
                     CDEEEKIKLTYKLSKASKAKIEVGRFKGLGEMMPAQLKETTMHPEKRSLLKVTLEDVQ
                     NVDKIVDDLMGKKPEKRFQFIYEQALVKMDQIINKLDI"
     gene            4122. .5447
                     /gene="RP228"
     CDS             4122. .5447
                     /gene="RP228"
                     /codon_start=1
                     /transl_table=11
                     /product="TAIL-SPECIFIC PROTEASE PRECURSOR (ctp)"
                     /protein_id="CAA14691.1"
                     /db_xref="GI:3860791"
                     /db_xref="GOA:Q9ZDU6"
                     /db_xref="InterPro:IPR001478"
                     /db_xref="InterPro:IPR004447"
                     /db_xref="InterPro:IPR005151"
                     /db_xref="UniProt/TrEMBL:Q9ZDU6"
                     /translation="MLLRFIIALFLSINCAIEGKETENKISNQEAYKQFQDVFERIEK
                     DYVQVPDRQKMIDEAINGMLNSLDPHSNYYTDEDLEDIFTFTKGEFGGIGVEIMYDSG
                     AIKIISSIDDLPAFKAGLKGGDYIVGVNDELVSTLGPNKAIKEMRGTPGTKVRLLIIK
                     EEEAKPQELELTREIVKIKPIKAHLEKNNIAYIRITTFNESTISELKAAVKKLKTESK
                     DNLKGIILDLRNNAGGILDQAIAVSDYFIDSGVIVTTKGRTTSSNSETKANEFSLKAP
                     KVPMIVLINGNSASASEIVAGALQDHKRAIILGTKSFGKGSVQALTQINSRAAVKLTI
                     SKYYTPSGRSIQAEGIEPDILIEPAKVEYPEVKKIDKRFSESSLKNYLKNDNAKNKDS
                     NKETKTKNNKQEESELSELYKKDYQFARAYDVITGLIINTNLETQGKAK"
     gene            5653. .7146
                     /gene="RP229"
     CDS             5653. .7146
                     /gene="RP229"
                     /codon_start=1
                     /transl_table=11
                     /product="HISTIDINE KINASE SENSOR PROTEIN (barA)"
                     /protein_id="CAA14692.1"
                     /db_xref="GI:3860792"
                     /db_xref="GOA:Q9ZDU5"
                     /db_xref="InterPro:IPR003594"
                     /db_xref="InterPro:IPR003661"
                     /db_xref="InterPro:IPR004358"
                     /db_xref="InterPro:IPR005467"
                     /db_xref="UniProt/TrEMBL:Q9ZDU5"
                     /translation="MNHNKYIVRLSFIILFLNIVINMLFYRYFMIKEMIIKQVALEHT
                     KIVELYTDNIWNIHKNVISKLHKFDYLKLLQDQDFINFVKITAQWFTNLNINISLYDL
                     KGNKFITSNMLHMYSVDNYKDDSLIEIVITKIDKIFLKSFTSAAPLRDAFKGITSHIL
```

```
                     LPKVVIENESDLTIEHASFVTSYIPVIDHNLDFPVDAVFEINTNITSQWKNITSLEQK
                     VFITPIIIFIIFCTIIISNTNYARQIIEEQLETNRNLKAQIEKVEKTSSSNTKFFANI
                     SHELRTPLNAIIGFSEILMSERDTEKSKNYIKDIHDAGKHLLSMINDILDLSKASADK
                     LKVDNIDLDLNKLISSSLILIKPRAYQAEVELISRLPKEHVVINADPKRLKQVLLNLL
                     SNAVKFTNSGGSVTISLEKDELAKLVYIKVIDTGIGIEEKDIPKTLSAFGQIDSELSR
                     KYEGTGLGLPLTKKLVELMNGKFDLQSKINKGTTVTITFKYDGSIEI"
gene                 7374. .8126
                     /gene="RP230"
CDS                  7374. .8126
                     /gene="RP230"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA14693.1"
                     /db_xref="GI:3860793"
                     /db_xref="GOA:Q9ZDU4"
                     /db_xref="InterPro:IPR001440"
                     /db_xref="UniProt/TrEMBL:Q9ZDU4"
                     /translation="MKLIILLITFLFSLFSLGKSEMPQGKPLKYAVNNDFENRLDEQE
                     QEIRKLIGKVEILQHKIDLLKQNLNMLNQEENIEVLETDDFKKQDIFDIALLEGMHDN
                     VSKKTFEVNKDIAPYKQAYDLALAAYKDNKLTEAKDKFKNFIQKYPNNSLISNAYFWY
                     AECFFKQKDYNGAAINYLKCYKESPKGAKSSDGLLKLALSLGELKKMQEACNILAKLD
                     KEFPINRTSVSKKMTEDAKIKFGCKINKNNKI"
gene                 8123. .8785
                     /gene="RP231"
CDS                  8123. .8785
                     /gene="RP231"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA14694.1"
                     /db_xref="GI:3860794"
                     /db_xref="GOA:Q9ZDU3"
                     /db_xref="UniProt/Swiss-Prot:Q9ZDU3"
                     /translation="MTDILDEVLSDQNEEKRLIFFKKLVPIIIIISIVVITIMVIIKN
                     NKDNRIKNNQKNGDILVKTIGLDTTKDNKALAFNTLENLITSSNTKIKEIAALEQVAI
                     RMSEKKYLGAKNLLNKIIDNEKYSEISTAYARIAWCCIVLDDQSLNIQDKGKLQKYLN
                     YFDDEKKPFWATATIIKAILDIKHNMKTQAEKNLKNLLASNNVSDLLKDQAKALLVSL
                     SK"
gene                 8812. .10059
                     /gene="RP232"
CDS                  8812. .10059
                     /gene="RP232"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="CAA14695.1"
                     /db_xref="GI:3860795"
                     /db_xref="GOA:Q9ZDU2"
                     /db_xref="InterPro:IPR002372"
                     /db_xref="InterPro:IPR011047"
                     /db_xref="UniProt/Swiss-Prot:Q9ZDU2"
                     /translation="MKKKLALLLLPFILISCNGLGSKSVKNIVDLTPKLVIQTNEPIY
                     LDSNTKIYPFNVNMLKNKQYSPAKSKMIAEPVFIGDMIYTLDIRANISAFSIEKNKII
                     WSYNLSKHKKDNYIGGGILHHNGKLYITYGARLLIVLDAKSGYEIIRKELPDIIRIKP
                     IALNDHTILVQTISNQTIALDSETLKTLWDHESIAEILSTSYSMIPIVQHDNVIVTYN
                     TGQVLALNIKNGEVKWNFEFTNLNDHTAIPNFDTSSILCTPVHDSMNLYIATGLGKLI
                     KLNLLTGSVLWQINADDIQSMSLIGNSLFIINNARQIAALNPETGKVKFVADLNYEKN
                     DKRLKSTTFLVPFVGVDNNNQRSLNVISVDGILYNFNIDSNGLKMNPHIIKIIKNIRY
                     YGLRSNNTLYFSTDRQVIFGSQI"
gene                 10279. .10746
                     /gene="RP233"
CDS                  10279. .10746
                     /gene="RP233"
                     /codon_start=1
                     /transl_table=11
                     /product="50S RIBOSOMAL PROTEIN L13 (rplM)"
                     /protein_id="CAA14696.1"
                     /db_xref="GI:3860796"
                     /db_xref="GOA:Q9ZDU1"
                     /db_xref="InterPro:IPR005822"
                     /db_xref="InterPro:IPR005823"
                     /db_xref="UniProt/Swiss-Prot:Q9ZDU1"
                     /translation="MKTYSAKPSEIEKKWWVIDAKNVVLGRLASRVAIMLRGKHKPSF
                     TPHLDCGDNIIIINAEHIKLTGKKLNPKDGKVYYRHTGFPGGIKDTTAGKILSGKYPE
                     RVIKMAVKRMITRNVLGAKQMSNLYVYANCEHPHMAQQPTIYDFASENPKNKK"
gene                 10753. .11238
                     /gene="RP234"
CDS                  10753. .11238

  Query Match              6.1%;  Score 59;  DB 1;  Length 312430;
  Best Local Similarity   49.8%;  Pred. No. 0.53;
  Matches  149;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  ||||| ||  |||| ||  ||||| |||| | |  ||    |||||
Db     263304 TAAAGATTTTGAACTTATTATTATTGATGATTGTTCAAAAGATAATAGTGCTCTGATAGC 263363

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
               ||||       ||| |      |  || ||         |||||||  | ||   | |
Db     263364 TAAGAAATATGCAGAGCAGGACCCACGTATTAAGATAATAATAAATGAGCATAATAAAAG 263423

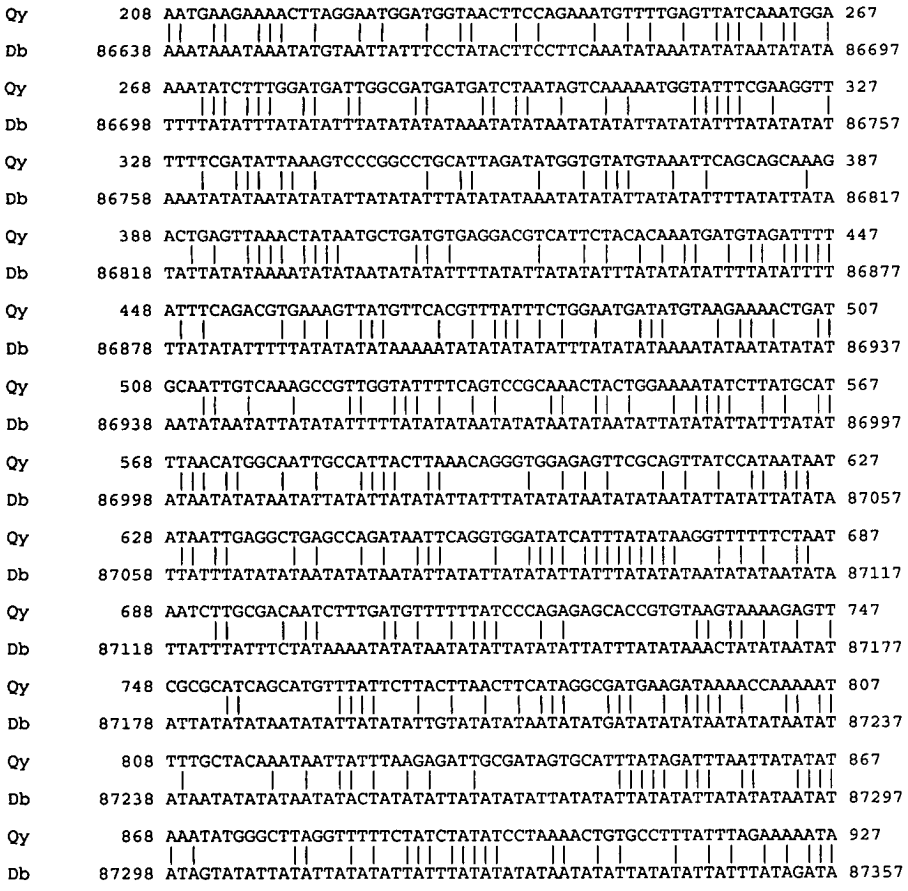
Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              | |   |  |  ||  | |   |||| |     | ||||| || ||| ||     ||
Db     263424 ATTATCAGCAAGTTTGAATATTGCTTTTAAAAAGGCTAATGGGAAGTATTTTACTTGGAC 263483

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |   |||||| ||||  |      || | | | |  | ||  |  ||  |||| |  |
Db     263484 TTCAGATGATAATCTTTTCCATGAAGAGGCTTTAGCAAAAATGGTTAAGATACTTGATAA 263543

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATA 403
                 ||||   ||| | | || |||    |||   |   ||   ||||  |    |||
Db     263544 TGCACCTGAGATAGGTCTTGTATATACTGATTATACTTTAATCGATGAGAAAGGTAATA 263602


RESULT 11
AP002761
LOCUS       AP002761              196988 bp    DNA     linear   PRI 11-JUL-2002
DEFINITION  Homo sapiens genomic DNA, chromosome 11 clone:RP11-800A3, complete
            sequence.
ACCESSION   AP002761
VERSION     AP002761.3  GI:21728155
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 196988)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Jul 10, 2002 this sequence version replaced gi:11611630.
FEATURES             Location/Qualifiers
     source          1. .196988
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-800A3"
ORIGIN

  Query Match              6.1%;  Score 58.8;  DB 8;  Length 196988;
  Best Local Similarity   42.5%;  Pred. No. 0.68;
  Matches  312;  Conservative    0;  Mismatches  422;  Indels    0;  Gaps    0;
```



```
Qy        208 AATGAAGAAAACTTAGGAATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGA 267
Db      86638 AAATAAATAAATATGTAATTATTTCCTATACTTCCTTCAAATATAAATATATAATATATA 86697

Qy        268 AAATATCTTTGGATGATTGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTT 327
Db      86698 TTTTATATTTATATATTTATATATATAAATATATAATATATATTATATATTTATATATAT 86757

Qy        328 TTTTCGATATTAAAGTCCCGGCCTGCATTAGATATGGTGTATGTAAATTCAGCAGCAAAG 387
Db      86758 AAATATATAATATATATTATATATTTATATATAAATATATATTATATATTTTATATTATA 86817

Qy        388 ACTGAGTTAAACTATAATGCTGATGTGAGGACGTCATTCTACACAAATGATGTAGATTTT 447
Db      86818 TATTATATAAAATATATAATATATATTTTATATTATATATTTATATATATTTTATATTTT 86877

Qy        448 ATTTCAGACGTGAAAGTTATGTTCACGTTTATTTCTGGAATGATATGTAAGAAAACTGAT 507
Db      86878 TTATATATTTTTATATATATAAAAATATATATATATTTATATATAAAATATAATATATAT 86937

Qy        508 GCAATTGTCAAAGCCGTTGGTATTTTCAGTCCGCAAACTACTGGAAAATATCTTATGCAT 567
Db      86938 AATATAATATTATATATTTTTATATATAATATATAATATAATATTATATATTATTTATAT 86997

Qy        568 TTAACATGGCAATTGCCATTACTTAAACAGGGTGGAGAGTTCGCAGTTATCCATAATAAT 627
Db      86998 ATAATATATAATATTATATTATATATTATTTATATATAATATATAATATTATATTATATA 87057

Qy        628 ATAATTGAGGCTGAGCCAGATAATTCAGGTGGATATCATTTATATAAGGTTTTTTCTAAT 687
Db      87058 TTATTTATATATAATATATAATATTATATTATATATTATTTATATATAATATATAATATA 87117

Qy        688 AATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGAGCACCGTGTAAGTAAAAGAGTT 747
Db      87118 TTATTTATTTCTATAAAATATATAATATATTATATATTATTTATATAAACTATATAATAT 87177

Qy        748 CGCGCATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAAT 807
Db      87178 ATTATATATAATATATTATATATTGTATATATAATATATGATATATATAATATATAATAT 87237

Qy        808 TTTGCTACAAATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATAT 867
Db      87238 ATAATATATATAATATACTATATATTATATATATTATATATTATATATTATATATAATAT 87297

Qy        868 AAATATGGGCTTAGGTTTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATA 927
Db      87298 ATAGTATATTATATTATATATTATTTATATATATAATATATTATATATTATTTATAGATA 87357

Qy        928 AAATATATTATAAA 941
Db      87358 TAATATATTATATA 87371
```

```
RESULT 12
YEU46859
LOCUS       YEU46859               26913 bp    DNA     linear   BCT 16-FEB-2001
DEFINITION  Yersinia enterocolitica lipopolysaccharide O-side chain
            biosynthesis genes.
ACCESSION   U46859 U18674 U25113
VERSION     U46859.1  GI:1197638
KEYWORDS    .
SOURCE      Yersinia enterocolitica (type 0:8)
  ORGANISM  Yersinia enterocolitica (type 0:8)
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Yersinia.
REFERENCE   1  (bases 15672 to 21224)
  AUTHORS   Zhang,L., Toivanen,P. and Skurnik,M.
  TITLE     The gene cluster directing O-antigen biosynthesis in Yersinia
            enterocolitica serotype 0:8: identification of the genes for
            mannose and galactose biosynthesis and the gene for the O-antigen
            polymerase
  JOURNAL   Microbiology 142 (Pt 2), 277-288 (1996)
    PUBMED  8932701
REFERENCE   2  (bases 1 to 26913)
  AUTHORS   Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C.,
            Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
  TITLE     Bacterial polysaccharide synthesis and gene nomenclature
  JOURNAL   Trends Microbiol. 4 (12), 495-503 (1996)
    PUBMED  9004408
REFERENCE   3  (sites)
  AUTHORS   Zhang,L., Radziejewska-Lebrecht,J., Krajewska-Pietrasik,D.,
            Toivanen,P. and Skurnik,M.
  TITLE     Molecular and chemical characterization of the lipopolysaccharide
            O-antigen and its role in the virulence of Yersinia enterocolitica
            serotype O:8
  JOURNAL   Mol. Microbiol. 23 (1), 63-76 (1997)
    PUBMED  9004221
REFERENCE   4  (sites)
  AUTHORS   Bengoechea,J.A. and Skurnik,M.
  TITLE     Temperature-regulated efflux pump/potassium antiporter system
            mediates resistance to cationic antimicrobial peptides in Yersinia
  JOURNAL   Mol. Microbiol. 37 (1), 67-80 (2000)
    PUBMED  10931306
REFERENCE   5  (bases 20640 to 26913)
  AUTHORS   Zhang,L., Booth,I., Toivanen,P. and Skurnik,M.
  TITLE     Genetic characterization of a novel locus of Yersinia
            enterocolitica serotype O:8 responsible for temperature regulation
            of lipopolysaccharide O-side chain biosynthesis
  JOURNAL   Unpublished
REFERENCE   6  (sites)
  AUTHORS   Bengoechea,J.A., Pinta,E., Oertelt,C., Holst,O.,
            Radziejewska-Lebrecht,J., Piotrowska-Seget,Z., Venho,R. and
            Skurnik,M.
  TITLE     Identification of the UDP-N-acetylglucosamine-4-epimerase (gne),
            chain length determinant (wzz) and O-antigen polymerase (wzy) genes
            of the O-antigen gene cluster of Yersinia enterocolitica serotype
            O:8
  JOURNAL   Unpublished
REFERENCE   7  (bases 1 to 26913)
  AUTHORS   Zhang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-1996) Lijuan Zhang, Department of Medical
            Microbiology, University of Turku, Kiinamyllynkatu 13, 20520-Turku,
            Finland
COMMENT     On or before Feb 20, 1996 this sequence version replaced
            gi:1098515, gi:806861.
FEATURES             Location/Qualifiers
     source          1. .26913
                     /organism="Yersinia enterocolitica (type 0:8)"
                     /mol_type="genomic DNA"
                     /strain="8081-c"
                     /db_xref="taxon:34054"
     gene            1. .187
                     /gene="adk"
     CDS             <1. .187
                     /gene="adk"
                     /codon_start=2
                     /transl_table=11
                     /product="adenylate kinase"
                     /protein_id="AAC60759.1"
                     /db_xref="GI:1197639"
                     /translation="VRKRLVECHQQTAPLVSYYRKEADAGNTQYFKLDGTRKVAEVSA
                     NWRLFSVNLDGKIVKAA"
     gene            259. .1230
                     /gene="hemH"
     CDS             259. .1230
                     /gene="hemH"
                     /function="ferrochelatase"
                     /codon_start=1
                     /transl_table=11
                     /product="HemH"
                     /protein_id="AAC60760.1"
                     /db_xref="GI:806863"
                     /translation="MKQSKLGVLMVNLEHRMPPTPQAVKRYLAEFLSVRRVVDTSPWL
                     WWPLLRGVILPIRSPRVAKLYQSVWMDEGSPLLVYSRRQQKALAERMPEIPVELGMSY
```

GSPNLPDAIDKLLAQGVTKLVVLPLYPQYSWSTSAAVWDAVARILASYRRLPSISFIR
DYAEHPAYISALKQSVENSFVQHGKPDRLVLSFHGIPKPYAQLGDDYPQRCEDTSRAL
RAEIALPAETKSRVHIQSRFGREPWLTPYTDETLKSLPSQGVKHIQLICPGFSADCLE
TLEEIKEQNREIFLHAGGEKFEYIPALNDDKGHIDLLEQLVRDHLSC"

misc_feature 1561. .1630
/function="JUMPstart sequence"

gene 1997. .2782
/gene="ddhA"
/note="synonym: rfbF; named in accordance with the
Bacterial Polysaccharide Gene Nomenclature scheme"

CDS 1997. .2782
/gene="ddhA"
/function="putative glucose-1-phosphate
cytidylyltransferase"
/codon_start=1
/transl_table=11
/product="DdhA"
/protein_id="AAC60761.1"
/db_xref="GI:1197640"
/translation="MECKVKAVILAGGLGTRLSEETVIKPKPMVEIGGMPILWHIMKL
YSSYGINDFIICCGYKGYVIKEYFANYFMHMSDITFSMRDNEMKVHQKRVEPWNVTLV
DTGEHSMTGGRLKRVKDYVKDDDAFCFTYGDGVSDINISELINFHKSHGKKATLTATY
PPGRFGALDIENKQVRSFKEKPKGDGALINGGYFVLSPKVIDLIDGDKSIWEQEPLMI
LADNGELMAYEHTGFWQPMDTLRDKIYLQQLWEEGQAPWKVWE"

gene 2788. .3864
/gene="ddhB"
/note="synonym: rfbG; named in accordance with the
Bacterial Polysaccharide Gene Nomenclature scheme"

CDS 2788. .3864
/gene="ddhB"
/function="putative CDP-glucose 4,6-dehydratase"
/codon_start=1
/transl_table=11
/product="DdhB"
/protein_id="AAC60762.1"
/db_xref="GI:1197641"
/translation="MINEKFWQDKRVFITGHTGFKGSWLSLWLQNMGAKVKGYALDAP
TKPSLFEAANVADGMISEVGDIRNYAQLLKSINDFKPEIVFHMAAQSLVRLSYLEPVE
TYSTNVMGTVYLLEALRHVGGAKAIVNVTSDKCYENKEWVWGYREDESMGGYDPYSNS
KGCAELVTSSYRNSFFNPVKYEQHGTALATQRAGNVIGGGDWALDRVVPDMLRAFENS
EPVIIRNNPYAIRPWQHVLEPLSGYLLLAQRLYIDGIKYAEGWNFGPDDTDSSQVQSI
VDKMVKYWGGDACWQLDDHEHPHEANYLKLDCSKAKVLLGWQSRWNLDCTLEYIVAWH
KAWLRGVNMHEYSIVEINKYMASK"

gene 3988. .4512
/gene="wbcA"
/note="synonym: orf4.0; named in accordance with the
Bacterial Polysaccharide Gene Nomenclature scheme"

CDS 3988. .4512
/gene="wbcA"
/function="putative 6-deoxy-gulose synthetase-1"
/codon_start=1
/transl_table=11
/product="WbcA"
/protein_id="AAC60763.1"
/db_xref="GI:1197642"
/translation="MIINITELNISGCYLIESPIFSDERGEFVKTHHQEIFKNFGLEI
PSAEEYYSRSKNNVIRGMHFQQYPDDHNKLVFCPEGEVLDVFLDIRKDSNTYGQFMSF
ISNPHNRRSIFLAKGIAHGFLSMKDNTLIVCKTSTVHSPSRDSGIHWNSFGFKWPVEN
PIISDKDRNLDCFF"

gene 4545. .5423
/gene="wbcB"
/note="synonym: orf4.5; named in accordance with the
Bacterial Polysaccharide Gene Nomenclature scheme"

CDS 4545. .5423
/gene="wbcB"
/function="putative 6-deoxy-gulose synthetase-2"
/codon_start=1
/transl_table=11
/product="WbcB"
/protein_id="AAC60764.1"
/db_xref="GI:1197643"
/translation="MRILITGASGYIGSNLACYLNSNGFEVFGLIRSEVVDENKIKLL
SGVKLEILDEKNLCSLVENINPDIVIHIASLTSVTHDYSTIENLLRSNIEFPTKLLEA
MEVAGVKKFINTGTTWQNYNSADYEPVNLYAATKQAFEDILKYYIFAKGFSSITLRLF
DTYGPNDTRKKLIPLLDRLAETKESLDMSEGNQEIELVHINDICSAYKTAILKLQDGQ
PGYANVYSVDTGSRMSLKKLVKLYEKVNDVTLNINWGARPPREREIMRLCTNLTVLPN
WSPTIPLDEGINKKNR"

gene 5642. .6493
/gene="wbcC"
/note="synonym: orf5.6; named in accordance with the
Bacterial Polysaccharide Gene Nomenclature scheme"

CDS 5642. .6493
/gene="wbcC"
/function="putative 6-deoxy-gulose transferase"
/codon_start=1
/transl_table=11
/product="WbcC"
/protein_id="AAC60765.1"
/db_xref="GI:1197644"
/translation="MINNWIKKTHISIVYNVNKTNLGADINFLLAPKLASGKYCWLFG
SDDKLFPGAIDYLEDFLELDNDIYLVERTEYDFNFAKVNRAGRRWMSTGSRIYNSNNK
NELLDYFSNSISIGAIFSFLSSNIVKRARWEAIEFDDNYIGSAYPHVYILLSILNEGA
VVNYINKPLVMCRGDNDHFSKDGVINRIELDFRGFLSFSDEFYSNQSQIKSKFNSILC
HERPFLKTIAVVSTIGTKEEKERIRGYYNKLGMSNIVIHFIYLLKPLLRLSKYIRDGF
LNWKIDK"

gene 6526. .7815
/gene="wzx"
/note="synonym: rfbX; named in accordance with the
Bacterial Polysaccharide Gene Nomenclature scheme"

CDS 6526. .7815
/gene="wzx"
/function="putative O-antigen 'flippase'"
/codon_start=1
/transl_table=11
/product="Wzx"
/protein_id="AAC60766.1"
/db_xref="GI:1197645"
/translation="MIKNTYEKILLNKSLNKALRNSGWLIFDKIASLTIGLWVGILLA
KYLGPANYGQIVYAVTYVTFFQTISILGLDSIVVRDLSNGVEKCSAIISTIFTARFLM
SMVCLIIAISLAYNSRPDKRLILSIISLNILFSVGSVFDYLVQSQSKSRLTICAKLIS

Query Match 6.0%; Score 57.6; DB 1; Length 26913;
Best Local Similarity 55.4%; Pred. No. 2.3;
Matches 155; Conservative 0; Mismatches 119; Indels 6; Gaps 2;

```
Qy        26 TTGCTATTCCGACATATAACCGCTCTTCATGTTTGGCTCGTTTACTTGATAGTATAATTC 85
             ||   |||||||| |||||  | ||         ||  ||   ||| | || |||||  ||
Db      5511 TTTGCATTCCGACCTATAATAGGTCAAGTCTATTATCTGAATTAATAGAAAGTATTGTTT 5570

Qy        86 AACAGGAGAACTATTGTCATGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAG 145
               ||    ||  || |  |||||    | ||| ||  | ||| |||||||||| ||    |
Db      5571 CTCAATGTAAGGATAGAAATGAT---ATAGAGATTTGTATTTCTGATAATGCATCTGATG 5627

Qy       146 ATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGATAAAATAAGAAAT---AGTACTTATC 202
             || | |||     |||    || | | |  |  | ||||      |     |    |||
Db      5628 ATAATACAATTGAAATGATTAACAATTGGATTAAGAAAACTCATATCTCGATAGTTTACA 5687

Qy       203 ATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAA 262
             || ||||| ||   ||| | ||   |||     ||||| | |   |  |  | ||||| ||
Db      5688 ATGTAAATAAAACTAACCTTGGGGCGGACATAAACTTTCTGTTAGCTCCTAAGTTAGCAT 5747

Qy       263 ATGGAAAATATCTTTGGATGATTGGCGATGATGATCTAAT 302
               || ||||||  |||| |  ||||    |||||||  | |
Db      5748 CCGGTAAATATTGTTGGTTATTTGGAAGTGATGATAAATT 5787
```

RESULT 13
AF461768
LOCUS AF461768 20511 bp DNA linear BCT 26-SEP-2002
DEFINITION Yersinia pseudotuberculosis serogroup IA O-antigen gene cluster,
complete sequence.
ACCESSION AF461768 U29692
VERSION AF461768.1 GI:23321093
KEYWORDS .
SOURCE Yersinia pseudotuberculosis
ORGANISM Yersinia pseudotuberculosis

```
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Yersinia.
REFERENCE   1  (bases 1 to 20511)
  AUTHORS   Hobbs,M. and Reeves,P.R.
  TITLE     Genetic organisation and evolution of Yersinia pseudotuberculosis
            3,6-dideoxyhexose biosynthetic genes
  JOURNAL   Biochim. Biophys. Acta 1245 (3), 273-277 (1995)
  PUBMED    8541300
REFERENCE   2  (bases 1 to 20511)
  AUTHORS   Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C.,
            Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
  TITLE     Bacterial polysaccharide synthesis and gene nomenclature
  JOURNAL   Trends Microbiol. 4 (12), 495-503 (1996)
  PUBMED    9004408
REFERENCE   3  (bases 1 to 20511)
  AUTHORS   Pacinelli,E., Wang,L. and Reeves,P.R.
  TITLE     Relationship of Yersinia pseudotuberculosis O antigens IA, IIA, and
            IVB: the IIA gene cluster was derived from that of IVB
  JOURNAL   Infect. Immun. 70 (6), 3271-3276 (2002)
  PUBMED    12011023
REFERENCE   4  (bases 1 to 20511)
  AUTHORS   Hobbs,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-1995) Department of Microbiology, University of
            Sydney, Building G08, Sydney, NSW 2006, Australia
REFERENCE   5  (bases 1 to 20511)
  AUTHORS   Pacinelli,E., Wang,L. and Reeves,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) Microbiology, University of Sydney, Sydney,
            NSW 2006, Australia
  REMARK    Nucleotide and amino acid sequences updated by submitter
COMMENT     On Sep 26, 2002 this sequence version replaced gi:1223963.
FEATURES             Location/Qualifiers
     source          1. .20511
                     /organism="Yersinia pseudotuberculosis"
                     /mol_type="genomic DNA"
                     /isolate="M444"
                     /db_xref="taxon:633"
                     /note="serogroup: IA"
     gene            207. .1196
                     /gene="ddhD"
     CDS             207. .1196
                     /gene="ddhD"
                     /note="DdhD"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-6-deoxy-delta-3,4-glucoseen reductase"
                     /protein_id="AAN23035.1"
                     /db_xref="GI:23321094"
                     /translation="MSLNVKLHPSGIIFTSDGTSTILDAALDSNIHIEYSCKDGTCGS
                     CKAILISGEVDSAENTFLTEEDVAKGAILTCCSKAKSDIELDVNYYPELSHIQKKTYP
                     CKLDSIEFVGEDIAILSLRLPPTAKIQYLAGQYIDLIINGQRRSYSIANAPGGNGNIE
                     LHVRKVVNGVFSNIIFNELKLQQLLRIEGPQGTFFVREDNLPIVFLAGGTGFAPVKSM
                     VEALINKNDQRQVHIYWGMPAGHNFYSDIANEWAIKHPNIHYVPVVSGDDSTWTGATG
                     FVHQAVLEDIPDLSLFTVYACGSLAMITAARNDFINHGLAENKFFSDAFVPSK"
     gene            1222. .2007
                     /gene="ddhA"
     CDS             1222. .2007
                     /gene="ddhA"
                     /note="DdhA"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose-1-P-cytidylyltransferase"
                     /protein_id="AAN23036.1"
                     /db_xref="GI:23321095"
                     /translation="MEIQVKAVILAGGLGTRLSEETVVKPKPMVEIGGKPILWHIMKL
                     YSSYGINDFVICCGYKGYVIKEYFANYFMHMSDITFCMRDNEMIVHQKRVEPWNVTLV
                     DTGEDSMTGGRLRRVKDYVKDDEAFCFTYGDGVSDVNIAELIEFHKSHGKQATLTATY
                     PPGRFGALDIKDKQVRSFKEKPKGDGALINGGYFVLSPKVIDLIDGDKSTWEQEPLMT
                     LAAQGELMAFEHAGFWQPMDTLRDKIYLHELWEEGRAPWKVWE"
     gene            1910. .3085
                     /gene="ddhB"
     CDS             1910. .3085
                     /gene="ddhB"
                     /note="DdhB"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-glucose-4,6-dehydratase"
                     /protein_id="AAN23037.1"
                     /db_xref="GI:23321096"
                     /translation="MPVSGSQWIRCVTRSICMNYGKKAGHLGRYGNNKMINNSFWQGK
                     RVFVTGHTGFKGGWLSLWLQTMGAKVKGYSLTPPTVPSLFETARVADGMQSEIGDIRD
                     QNKLLESIREFQPEIVFHMAAQPLVRLSYSEPVETYSTNVMGTVYLLEAIRHVGGVKA
                     VVNITSDKCYDNKEWIWGYRENEAMGGYDPYSNSKGCAELVTSSYRNSFFNPANYGQH
                     GTAVATVRAGNVIGGGDWALDRIVPDILRAFEQSQPVIIRNPHAIRPWQHVLEPLSGY
                     LLLAQKLYTDGAEYAEGWNFGPNDADATPVKNIVEQMVKYWGEGASWQLDGNAHPHEA
                     HYLKLDCSKAKMQLGWHPRWNLNTTLEYIVGWHKNWLSGTDMHEYSITEINNYMNTK"
gene                 3103. .4416
                     /gene="ddhC"
CDS                  3103. .4416
                     /gene="ddhC"
                     /note="DdhC; formerly known as rfbH"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-4-keto-6-deoxy-D-glucose-3-dehydrase"
                     /protein_id="AAN23050.1"
                     /db_xref="GI:23321109"
                     /translation="MSQEELRQQIAELVAQYAETAMAPKPFEAGKSVVPPSGKVIGTK
                     ELQLMVEASLDGWLTTGRFNDAFEKKLGEYLGVPYVLTTTSGSSANLLALTALTSPKL
                     GVRALKPGDEVITVAAGFPTTVNPTIQNGLIPVFVDVDIPTYNVNASLIEAAVSDKTK
                     AIMIAHTLGNLFDLAEVRRVADKYNLWLIEDCCDALGSTYDGKMAGTFGDIGTVSFYP
                     AHHITMGEGGAVFTQSAELKSIIESFRDWGRDCYCAPGCDNTCKKRFGQQLGSLPFGY
                     DHKYTYSHLGYNLKITDMQAACGLAQLERIEEFVEKRKANFKYLKDALQSCADFLELP
                     EATENSDPSWFGFPITLKEDSGVSRIDLVKFLDEAKVGTRLLFAGNLTRQPYFHDVKY
                     RVVGELTNTDRIMNQTFWIGIYPGLTHDHLDYVVSKFEEFFGLNF"
repeat_region        4427. .4442
                     /rpt_type=inverted
gene                 4453. .5310
                     /gene="prt"
CDS                  4453. .5310
                     /gene="prt"
                     /note="Prt; formerly known as rfbS"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-paratose synthetase"
                     /protein_id="AAB48322.1"
                     /db_xref="GI:1223965"
                     /translation="MKILITGVSGYLGSQLANALMLEHEVAGTVRAGSVCNRITDIGN
                     VNLINVTDSGWIDKVLSFSPDVVINTVALYGRKGELLSELVDANIQFPLRILEMLVST
                     GKGLFFQCGTSLPADVSQYALTKNQFTELAREYCNKFSGKFIELKLEHFFGPFDDSTK
                     FTTYVINSCRSHSDLKLTAGLQRRDFIYINDLINAFKIMISKSESLISGESISIGSGH
                     AVTIKEFVETVAKMTSYQGNLQFGAIPTRENELMYSCASLARIQELGWLCQYSLNSAI
                     KDTLNRMRV"
gene                 5311. .6594
                     /gene="wbyH"
CDS                  5311. .6594
                     /gene="wbyH"
                     /codon_start=1
                     /transl_table=11
                     /product="WbyH"
                     /protein_id="AAN23051.1"
                     /db_xref="GI:23321110"
                     /translation="MDMKFCKNRYINTVILGAGVAGLGAALKAREVNIEYEIFEAKTS
                     AGGLLDNFVIDGYRFDNAVHLSFATEEKVRDVFDRTPYFSHSPDAVNFETDKWLKHPV
                     QNNLYPLPAIEKVTLIESLIHRPESELADNYESWLIHQYGNEIAKRYPMRYTRKYWQC
                     NANELSVSWIGNRLRRAKLSEILYGAFTDDTPNDYYAKEMRYPKKGGFKSFIQPLIDE
                     SIINYNYEVTEIDLSGKTIKFSNGVIISYENLINTLPLPLFIEFCGANVPTSVKKSAS
                     KLGATSLDLISVGFNKSIIKDLWFYIYDEDIYASRAYSPSVKSSDNVPDGCSSLQFEI
                     YNPGRDSQYTVQELKDNVIYAMKKMGVANNEDIVFLHHKRLPWGNVVFKKDMEKYRDI
                     VREYIFDNQVKSVGRFGEWDYLWSNQSFLSGYSSI"
gene                 6776. .8119
                     /gene="wzx"
CDS                  6776. .8119
                     /gene="wzx"
                     /note="Wzx"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="O-unit flippase-like protein"
                     /protein_id="AAN23038.1"
                     /db_xref="GI:23321097"
                     /translation="MRINYQMVWGLLNQGCNIFIPFVIMFMSYKYLPLEMASVWVIFL
                     SMISLITLFDFGLSPAIVRNVSYVISGAQNLVKRGIDDIIIKDVISYPLLSRLLFDIK
                     RIYLYLSLIAFFIIVIGGLWYFYYISPLDIKIEVAYSWVLFSSALIINLIYLYYVPVL
                     TGLGEIENSYKANVFGRIIWFFLSLLVFIFSPSILLLSLSFVFSVFINRYLCYFFYIR
                     NSHVKAIDKTEMDKKSTVPFIGHNAAKLGVVSLGSFLINRSTTLIVGIVCPIVTAGQF
                     VLSMQVFFALMAISNILLSIKIPEISQAVAKREHYTVKILIYKVVAFSSSIYLLGFVT
                     FLILAEYVLPTIGVSLSFLPLDYLLILGLIYFLEMNHSICATIITSKNNVPFVFASLF
                     SGGLIILLSIILSVHFKYGVVGLILAQGIVQLMYNNWKWPLMVYSEFIKNNIEDR"
     gene            8130. .9122
                     /gene="wbyM"
     CDS             8130. .9122
                     /gene="wbyM"
                     /note="WbyM"
                     /codon_start=1
                     /transl_table=11
                     /product="putative paratose transferase"
                     /protein_id="AAN23039.1"
                     /db_xref="GI:23321098"
                     /translation="MKSNIIIDNSNSDDVLLTIAIPTYKRFGLLKETLKSVFSLEFNI
                     PVEVIVVDNDPEHPELVLSEMAEFERYDFIYYKNIENYGMFGNWNQCLNLGRGKLITI
                     LHDDDLLCANFPRELDRFLKEYSCHDDIPLIGFGHYLLDQRKSEDKVNKNILYKISKK
                     IFNAFRKYIIKPKVVSITLDEIFWGNIFCGTLGVVMDRNKALSISGFDEKNYPVSDYD
                     FWIRWIQKYGPIKYNKTKVSYYRILDNESMRPDVISNFIKKTYELRLMMIKETPLLVN
                     QESNAINLKKLDEYSFNVNWSREGDFNLGILDTVNYIYIKIKCISCKYFSFWLH"
     gene            10136. .11245
                     /gene="wbyB"
     CDS             10136. .11245
                     /gene="wbyB"
                     /note="WbyB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative 6-deoxy-D-mannoheptose transferase"
                     /protein_id="AAN23040.1"
                     /db_xref="GI:23321099"
                     /translation="MNILIVYLGRKGGGAHYTYELVNELSKKTNVSLLVSKQAENIKS
                     YESLNISLNTINTFNGMFSFILSSFKLPFIKNKIKKIILDEKIDIVISTMTHVWNGFI
                     FTPNFMMKIPYILTIHDAVPHSGEGNFILNLLNNKDIKNANAYIALTEHVKNQFINIY
                     GDKKDICILPLGSFNFKSGEIKSWCGVRPLRILFFGRIHEYKGIDLLLDSMIILQEKY
                     DLELSIYGQGDLSDYVDKIDVIKKINIENRWIGDSEVGDILQAHDVCVIPYKDASQSG
                     VIPVAQSAGLPTIVNPVDGLIEQIEKDKTSLVTDRVDSVSLALSIEKIIKNPELVNKL
                     SIGSLEYASRKLSWVPIVNKLITYSKETLNSSDRS"
     gene            11325. .12410

  Query Match           5.9%;  Score 57;  DB 1;  Length 20511;
  Best Local Similarity  50.8%;  Pred. No. 3.2;
  Matches  166;  Conservative    0;  Mismatches  155;  Indels    6;  Gaps    1;

Qy        15 TCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTGGCTCGTTTACTTGA 74
             ||| || ||||| || || || |||||||| | | |   | ||          |||  |
Db      8174 TCTACTGACTATAGCAATACCCACATATAAGAGATTTGGTTTGTTAAAGGAAACACTAAA 8233

Qy        75 TAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTTATTGTTTGTGATAA 134
             ||| || ||  || ||      | | | ||  |  | | || || |||||   ||||||
Db      8234 AAGTGTATTTTCACTGG------AATTTAATATTCCAGTAGAAGTAATTGTCGTTGATAA 8287

Qy       135 TGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGATAAAATAAGAAATAG 194
             | | |   | ||      |  |    || |  | | |        | ||        ||
Db      8288 CGATCCTGAACATCCTGAATTAGTGTTATCAGAAATGGCAGAGTTTGAACGTTATGATTT 8347

Qy       195 TACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTCCAGAAATGTTTTGA 254
             ||  ||| ||  |||| ||||||||  || |||  || || |  |  |||| ||  |
Db      8348 TATATATTATAAAAATATAGAAAACTATGGGATGTTCGGCAATTGGAATCAATGCTTGAA 8407

Qy       255 GTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTAATAGTCAAAAATGG 314
             ||    |  || || |   ||   ||  |    ||||||||||| ||      ||
Db      8408 TTTGGGGAGAGGGAAGTTAATTACTATATTACATGATGATGATCTATTATGTGCTAACTT 8467

Qy       315 TATTTCGAAGGTTTTTTCGATATTAAA 341
             |  |    || |   |  | | |||||
```

```
Db      8468 TCCTCGTGAGTTAGATCGGTTCTTAAA 8494


RESULT 14
AL929058
LOCUS       AL929058              106063 bp    DNA     linear   VRT 02-DEC-2004
DEFINITION  Zebrafish DNA sequence from clone DKEY-69H9 in linkage group 20
            Contains the gene for a novel protein similar to human and mouse
            RALBP1 associated Eps domain containing 1 (REPS1), the 5' end of
            the gene for a novel protein similar to human and mouse cytochrome
            P450, family 39, subfamily A, polypeptide 1 (CYP39A1) and two CpG
            islands, complete sequence.
ACCESSION   AL929058
VERSION     AL929058.4  GI:25136754
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 106063)
  AUTHORS   Beasley,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Nov 19, 2002 this sequence version replaced gi:24817085.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep DKEY-69H9 is
            from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            Clone-derived Zebrafish pUC subclones occasionally display
            inconsistency over the length of mononucleotide A/T runs and
            conserved TA repeats. Where this is found the longest good quality
            representation will be submitted.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            Any regions longer than 1kb tagged as misc-feature 'unsure' are
            part of a tandem repeat of more than 10kb in length where it has
            not been possible to anchor the base differences between repeat
            copies.  The region has been built up based on the repeat element
            to match the total size of repeat indicated by restriction digest,
            but repeat copies may not be in the correct order and the usual
            finishing criteria may not apply.
FEATURES             Location/Qualifiers
     source          1. .106063
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="20"
                     /clone="DKEY-69H9"
                     /clone_lib="DanioKey"
     gene            complement(68923. .96727)
                     /locus_tag="DKEY-69H9.1-001"
     mRNA            complement(join(68923. .68991,71435. .71522,78583. .78671,
                     83101. .83196,87232. .87365,89787. .89877,90069. .90164,
                     91108. .91201,92107. .92256,93269. .93443,93647. .93782,
                     96551. .96727))
```

```
                     /locus_tag="DKEY-69H9.1-001"
                     /product="novel protein similar to human and mouse
                     cytochrome P450, family 39, subfamily A, polypeptide 1
                     (CYP39A1)"
     CDS             complement(join(68923. .68991,71435. .71522,78583. .78671,
                     83101. .83196,87232. .87365,89787. .89877,90069. .90164,
                     91108. .91201,92107. .92256,93269. .93443,93647. .93782,
                     96551. .96727))
                     /locus_tag="DKEY-69H9.1-001"
                     /standard_name="OTTDARP00000005949"
                     /note="match: proteins: Sw:Q9JKJ9 Sw:Q9NYL5 Tr:BAC27530
                     Tr:Q8CFY8"
                     /codon_start=1
                     /product="novel protein similar to human and mouse
                     cytochrome P450, family 39, subfamily A, polypeptide 1
                     (CYP39A1)"
                     /protein_id="CAH69101.1"
                     /db_xref="GI:55251024"
                     /translation="MEWITLILTVSVVIISFHLLFGKSHPNAPPCIRGWIPWFGAAFE
                     FGKAPLHFIQQARAKYGPVFTVVAAGKRMTFVTLNEDFRVFFTSKDVDFEQAVQEPVH
                     NTASISKDNFFESHPTCSAIIKGRLTPGNTAMLSPHLCEEFNDHLESLGSEGSGQLNE
                     LIKSVMYPSVMSNLLGRCNSPSSALSRQEFLEKFTTYDEGFEYGSQLPEMFLKEWSNS
                     KHWLLSLLRKMVIKSEETLYSESDRKTLLQHLAASISEQYLPNYGLLLLWASLANAIP
                     VTFWAVAFILSNPTAYKIVMDQINSVLGRQDKQKTKVTLDDLQQMPYVKWCIMEAIRL
                     RAPGAITRKVVRPLKLQNYVIPPGDMLMLSPYWAHRNPKYFPDPEDFKPERWETADLE
                     KNVLLEGFVAFGGGKNQCPGRWYAIMELHMFVALILYKFEFAQLDPMPKPSPLHLVGT
                     QQPDGPCTVRYKHR"
     gene            join(100468. .100636,AL845306.9:2110. .29974)
                     /locus_tag="DKEY-217M5.1-001"
     mRNA            join(100468. .100636,AL845306.9:2110. .2233,
                     AL845306.9:3603. .3787,AL845306.9:6777. .6933,
                     AL845306.9:8286. .8407,AL845306.9:10699. .10864,
                     AL845306.9:10959. .11022,AL845306.9:11244. .11398,
                     AL845306.9:15168. .15289,AL845306.9:19471. .19551,
                     AL845306.9:19894. .19981,AL845306.9:23629. .23727,
                     AL845306.9:23822. .23894,AL845306.9:24058. .24188,
                     AL845306.9:25407. .25471,AL845306.9:25570. .25743,
                     AL845306.9:28117. .28206,AL845306.9:28287. .28435,
                     AL845306.9:29173. .29278,AL845306.9:29378. .29974)
                     /locus_tag="DKEY-217M5.1-001"
                     /product="novel protein similar to human and mouse RALBP1
                     associated Eps domain containing 1 (REPS1)"
                     /note="match: ESTs: Em:AL921008.1 Em:AW419923.1
                     Em:BG729384.1 Em:BI889436.1 Em:CB353398.1 Em:CD594178.1"
     CDS             join(100484. .100636,AL845306.9:2110. .2233,
                     AL845306.9:3603. .3787,AL845306.9:6777. .6933,
                     AL845306.9:8286. .8407,AL845306.9:10699. .10864,
                     AL845306.9:10959. .11022,AL845306.9:11244. .11398,
                     AL845306.9:15168. .15289,AL845306.9:19471. .19551,
                     AL845306.9:19894. .19981,AL845306.9:23629. .23727,
                     AL845306.9:23822. .23894,AL845306.9:24058. .24188,
                     AL845306.9:25407. .25471,AL845306.9:25570. .25743,
                     AL845306.9:28117. .28206,AL845306.9:28287. .28435,
                     AL845306.9:29173. .29278,AL845306.9:29378. .29446)
                     /locus_tag="DKEY-217M5.1-001"
                     /standard_name="OTTDARP00000005948"
                     /note="match: proteins: Sw:O54916 Sw:Q96D71"
                     /codon_start=1
                     /product="novel protein similar to human and mouse RALBP1
                     associated Eps domain containing 1 (REPS1)"
                     /protein_id="CAH69102.1"
                     /db_xref="GI:55251025"
                     /translation="MESLTLSDVEQKYYSELFLSCDVDNAKKVASNGKVLELFRAAQL
                     SNEVVLQITELCGATRLGHFGRSQFYIALKLIALAQSGLPLRLESLNTVRDLPLPRFV
                     MGKNEQETRHMVMFTDAENQGPFLPRPPGRVQAKKVSAHEMIQPCVPTADPQPDTTSP
                     VVSPHQSPPTSPLAWRKHKRQASGGNVDRQPSVPGVVWPHFREAQPGPVPGDGMWSAH
                     SPPPVQESWVSFTDTPPSSTLPMHPPSAQTESSTVRTVPSAMTTNEIQRQTSGYDDPW
                     KITDEQRQYYINQFKTIQADLTGLIPGSAAKEFFTKSKLPILELSHIWELSDFDKDGA
                     LTLDEFCAAFHLVVARKNGYDLPEKLPESLMPKLIDLDDSAGVPEPAPEVGFSASPVE
                     VTPNKSPSMPSLNQNWPELNQSNEQWETFSERSSSSQTLTQFDSNIAPADPDTAIVHP
                     VPIRMTPSKIHMQEMELKRTGSDHTHPTSPLMTKPPELSDETKLAAAIKFPGTAVGDG
                     YSSSDSFTSDQEPISSAVNRQRSHSGTSPEGLKVVAPPPPPRPHASHSRSSSLDMNR
                     NFAAVPAGSQQAGVVAFPPAVPPRPLPTQTSVPHGHRSVEGDGLPAHTSTSPQQMPEQ
                     PNFADFSQFKAFAVDQPADDGENPSDGGQAERLPEAAGTMRTAKNDVQAEERNATTVN
                     SAKVSTPLAPPPKPVRRRLKSEDELRPDVDDLPQKSNVIATVLATQPSIPRSLGKDKK
                     AIQASIRRNKETNTVLARLNSELQQQLKDLLEERISLEVQLEQLRPFSHL"
     misc_feature    102064
                     /note="Clone_left_end: DKEY-217M5"
ORIGIN

  Query Match             5.9%;  Score 56.4;  DB 5;  Length 106063;
  Best Local Similarity   46.8%;  Pred. No. 2.1;
  Matches  247;  Conservative    0;  Mismatches  276;  Indels    5;  Gaps    2;

Qy        431 CAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATTTCTGGAATGA 490
              || |  |  |||||||| |||||    | ||  ||| |||     ||   | ||   | |
Db      38758 CACAGCAATTAGATTTTTTTTCATGACTTAATTTTAAGTTTTTACTTGACTTTGTGGTTA 38817

Qy        491 TATGTAAGAAAACTGATGCAATTG----TCAAAGCCGTTGGTATTTTCAGTCCGCAAACT 546
              ||||| | | ||||||| || |||    |||        |   ||||      |   ||
Db      38818 TATGTCATATAACTGATTCAGTTGATGATCATTTAAAATTTAATTTAAGAAACTATAAAC 38877

Qy        547 ACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGTGGAGAG 606
                |  ||||||    ||  | | ||       ||         | | |   |  ||| ||
Db      38878 GATAAAAAATAAACAATAAAATAAATGACTAAAACAAATAAAATAATTAAATATGGTGAT 38937

Qy        607 TTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCA-GATAATTCAGGTGGATATCA 665
              |  | |    || ||||||  | | | | | |   |||||      |    |   |
Db      38938 GATGATGATGATGATGATAATAATAATAATGATAAATAATAATAATAATTATTAATAAATA 38997

Qy        666 TTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGA 725
              | |||  |  |  |    | ||| | ||   |  ||| ||  | || ||  |   |   |
Db      38998 ATAATAATAAATAATAATAATTATTATTATTATTATCATTATTATTGTTGTTGTTATGTA 39057

Qy        726 GCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCATAGG 785
              |  | | | ||||| ||||      | |  ||  || || | | | ||  | ||
Db      39058 TTTCTTTTTTATTAAAATAGTTGTTTACTTATTATTATTTTTTTAAATAAAAATAATTAT 39117

Qy        786 CGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGATAGTGC 845
              || |  | |  |    |||||  || |  || || |||||     |      |   |
Db      39118 TTATTATTACATTATTTAAAATAATGTTTTAATTAGTTATTAGTTAATCATTTAATATAT 39177

Qy        846 ATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCTAAAAC 905
              | | ||||   |||| ||| |  ||     | |   ||  |  ||  |  |  ||| |
Db      39178 AGTAATAGTAATAATAATAATTTTATTAATACGTTTTTTCATTAATTAAGTTTTTAATAA 39237

Qy        906 TGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAAT 953
              |     |    ||  || |  ||  |||||  || |    || |||
Db      39238 TAATAATAATAATAATAATAATAATAATATTAATAATAATAATTGAAT 39285


RESULT 15
CP000020_01
WPCOMMENT
  Sequence split into 29 fragments  LOCUS CP000020 Accession CP000020
     Fragment Name      Begin     End
     CP000020_00            1    110000
     CP000020_01       100001    210000
     CP000020_02       200001    310000
     CP000020_03       300001    410000
     CP000020_04       400001    510000
     CP000020_05       500001    610000
     CP000020_06       600001    710000
     CP000020_07       700001    810000
     CP000020_08       800001    910000
     CP000020_09       900001   1010000
     CP000020_10      1000001   1110000
     CP000020_11      1100001   1210000
     CP000020_12      1200001   1310000
     CP000020_13      1300001   1410000
     CP000020_14      1400001   1510000
     CP000020_15      1500001   1610000
     CP000020_16      1600001   1710000
     CP000020_17      1700001   1810000
     CP000020_18      1800001   1910000
     CP000020_19      1900001   2010000
```

```
     CP000020_20       2000001   2110000
     CP000020_21       2100001   2210000
     CP000020_22       2200001   2310000
     CP000020_23       2300001   2410000
     CP000020_24       2400001   2510000
     CP000020_25       2500001   2610000
     CP000020_26       2600001   2710000
     CP000020_27       2700001   2810000
     CP000020_28       2800001   2906179
Continuation (2 of 29) of CP000020 from base 100001 (CP000020 Vibrio fischeri ES114 chro

  Query Match             5.8%;  Score 56;  DB 1;  Length 110000;
  Best Local Similarity   51.7%;  Pred. No. 2.4;
  Matches  181;  Conservative    0;  Mismatches  160;  Indels    9;  Gaps    2;

Qy          1 ATGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTG 60
              || ||| ||       |||||||||||| ||||| || || | ||| ||  ||  |    ||
Db     108616 ATAAGTGAATTTATGCTTCTTACTATAGCTATACCAACTTTTAATCGAGCTGAAAAGCTG 108675

Qy         61 GCTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCA---TGATGAACTCGAG 117
              |      |||| | | |||   | | ||     |   |     |    |||||  ||
Db     108676 GACTTGCAACTTAAAAATATTTGTGACCAATTAATTGAACTAAAAAGAGATGATATCTGT 108735

Qy        118 GTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTA 177
              |  | |||   |||||   ||||   ||| ||||| ||   ||    || |       |
Db     108736 ATCTTAGTTCAAGATAACCATTCAGATGATCAAACAACAGAGATTAGTAATAAATATATG 108795

Qy        178 GATAAAATAAGAAATAG------TACTTATCATCTAAATGAAGAAAACTTAGGAATGGAT 231
              ||||  |   || ||       |  ||||  |  ||||||    || |||||  | |||
Db     108796 TCTAAATTTGAAAGTATAGACTTTGTTTATTTTAGAAATGAGTCTAATTTAGGCTTTGAT 108855

Qy        232 GGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGAT 291
              | ||  |  |    |||||       |||||  | ||||||  ||||||  |    |||
Db     108856 CGAAATGTTAATTTGTGTTTGACTAATTCAAAAAGCAAATATACTTGGATATTATCTGAT 108915

Qy        292 GATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAA 341
              ||||||    |    |  ||||  || |  |   || | ||   | || |
Db     108916 GATGATGATCTCAATAGTAATGCAATATTAACTATTATATCTGAAATAGA 108965


Search completed: May 11, 2006, 19:14:20
Job time : 3472.28 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 07:09:59 ; Search time 438.793 Seconds
(without alignments)
14626.705 Million cell updates/sec

Title: US-09-423-093B-3_COPY_2352_3314
Perfect score: 963
Sequence: 1 atgagtaaaaaacctcttct.....cggttttaatgcggaaataa 963

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	963	100.0	12441	2	AAX06750	Aax06750 Salmonell
c	2	55.4	5.8	344	4	AAH93307	Aah93307 Plasmodiu
c	3	54.8	5.7	2000	11	ACL35887	Acl35887 Rice stre
	4	53	5.5	10652	6	ABN80057	Abn80057 Human che
	5	50.6	5.3	906	3	AAA53723	Aaa53723 Campyloba
	6	50.6	5.3	906	8	ABT13678	Abt13678 Campyloba
c	7	50.6	5.3	11474	3	AAA53720	Aaa53720 LPS core
c	8	50.6	5.3	11474	8	ABT13665	Abt13665 Campyloba
	9	49.6	5.2	1524	14	ADZ70897	Adz70897 Human mat
	10	49.6	5.2	4660	14	ADZ70902	Adz70902 Human mat
	11	48.4	5.0	5654	4	AAS46623	Aas46623 Tumour su
	12	48.4	5.0	5654	6	ABL33874	Abl33874 Human imm
	13	48	5.0	6775	6	ABQ67159	Abq67159 Human ang
	14	47.4	4.9	2000	11	ACL37108	Acl37108 Rice stre
c	15	47	4.9	16258	6	ABK40038	Abk40038 Human che
c	16	47	4.9	16258	6	ABL70376	Abl70376 Chemicall
	17	47	4.9	22080	2	AAX06751	Aax06751 Salmonell
	18	46.8	4.9	276	8	ABX44966	Abx44966 Bovine ES
	19	46.8	4.9	933	8	ACA22926	Aca22926 Prokaryot
c	20	46.6	4.8	2000	11	ACL35363	Acl35363 Rice stre
	21	46.6	4.8	2501	14	ADV98013	Adv98013 Bisulfite
	22	46.6	4.8	2501	14	ADV98107	Adv98107 Bisulfite
c	23	46.6	4.8	5908	6	ABL32596	Abl32596 Human imm
	24	46.4	4.8	52242	9	ADA02666	Ada02666 Human MDM
	25	46.4	4.8	52242	10	ADB72404	Adb72404 Human MDM
	26	46.4	4.8	52242	10	ADE95914	Ade95914 Human MDM
	27	46.2	4.8	1011	8	ACA27781	Aca27781 Prokaryot
	28	46.2	4.8	1038	13	ADU25587	Adu25587 cDNA enco
c	29	46.2	4.8	2000	11	ACL37108	Acl37108 Rice stre
c	30	46.2	4.8	2131	14	ADZ71009	Adz71009 Human chr
	31	46.2	4.8	33053	6	ABQ67006	Abq67006 Human ang
	32	46	4.8	1356	10	ADF02944	Adf02944 Bacterial
	33	46	4.8	2482	14	ADZ71091	Adz71091 Human chr
c	34	45.4	4.7	5273	6	ABL32875	Abl32875 Human imm
c	35	45.4	4.7	8392	6	ABL33491	Abl33491 Human imm
	36	45.2	4.7	366	4	AAH93356	Aah93356 Human chr
c	37	45	4.7	5818	4	AAS46635	Aas46635 Tumour su
c	38	45	4.7	8896	6	ABN80327	Abn80327 Human che
	39	45	4.7	110000	13	ADS99457_1	Continuation (2 of
	40	44.8	4.7	784	14	ADZ71064	Adz71064 Human chr
c	41	44.8	4.7	804	6	AAS62093	Aas62093 Porcine m
	42	44.8	4.7	9347	6	ABL49336	Abl49336 Human pol
	43	44.8	4.7	16217	6	ABL32625	Abl32625 Human imm
	44	44.8	4.7	225587	12	ADQ97397	Adq97397 Human can
	45	44.6	4.6	1005	6	ABN67437	Abn67437 Streptoco

ALIGNMENTS

RESULT 1
AAX06750
ID AAX06750 standard; DNA; 12441 BP.
XX
AC AAX06750;
XX
DT 26-APR-1999 (first entry)
XX
DE Salmonella enterica serovar C2 O antigen gene cluster.
XX
KW O antigen; diagnosis; wbaR gene; wzx gene; wbaL gene; wbaQ gene;
KW wxy gene; wbaW gene; wbaZ gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Salmonella enterica.
XX
FH Key Location/Qualifiers
FT CDS 138. .1019
FT /*tag= a
FT /note= "abe gene; encodes AAW88324"
FT CDS 1019. .2359
FT /*tag= b
FT /product= "flippase"
FT /note= "wzx gene, this region is specifically claimed in
FT Claim 9; encodes AAW88325"
FT CDS 2362. .3314
FT /*tag= c
FT /product= "abequosyl_transferase"
FT /note= "wbaR gene, this region is specifically claimed in
FT Claim 9; encodes AAW88326"
FT CDS 3361. .3975
FT /*tag= d
FT /product= "acetyl_transferase"
FT /note= "wbaL gene, this region is specifically claimed in
FT Claim 9; encodes AAW88327"
FT CDS 3977. .5020
FT /*tag= e
FT /product= "rhamnosyl_transferase"
FT /note= "wbaQ gene, this region is specifically claimed in
FT Claim 9; encodes AAW88328"
FT CDS 5114. .6313
FT /*tag= f

```
FT                   /product= "O_antigen_polymerase"
FT                   /note= "wzy gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88329"
FT   CDS             6313. .7323
FT                   /*tag=  g
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88330"
FT   CDS             7310. .8467
FT                   /*tag=  h
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88331"
FT   CDS             8482. .9903
FT                   /*tag=  i
FT                   /note= "manC gene; encodes AAW88332"
FT   CDS             9911. .11347
FT                   /*tag=  j
FT                   /note= "manB gene; encodes AAW88333"
FT   CDS             11426. .12441
FT                   /*tag=  k
FT                   /note= "wbaP gene; encodes AAW88334"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88324, AAW88325, AAW88326, AAW88327, AAW88328, AAW88329,
DR   AAW88330, AAW88331, AAW88332, AAW88333, AAW88334.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 9; Fig 9; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Salmonella enterica serotype C2 (strain M67) O
CC   antigen. 11 Open reading frames within the gene cluster were identified
CC   on the basis of homology to known sequences. The genes encode proteins
CC   (see AAW88323-34) involved the synthesis of sugars present in the
CC   polysaccharide antigen, and in the transport or processing of
CC   polysaccharide or oligosaccharide units. The use of nucleic acid
CC   molecules derived from assembly and transport genes, particularly wbd
CC   (transferase), wzx (flippase) and wzy (polymerase) genes, within O
CC   antigen gene clusters improves the specificity of methods for the
CC   detection and identification of O antigens, e.g. in testing food- or
CC   faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of S. enterica
XX
SQ   Sequence 12441 BP; 3887 A; 1876 C; 2518 G; 4160 T; 0 U; 0 Other;
```

Query Match 100.0%; Score 963; DB 2; Length 12441;
Best Local Similarity 100.0%; Pred. No. 1.2e-192;
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2352 ATGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTG 2411

Qy         61 GCTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2412 GCTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTT 2471

Qy        121 ATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2472 ATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGAT 2531

Qy        181 AAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2532 AAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTC 2591

Qy        241 CAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2592 CAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTA 2651

Qy        301 ATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2652 ATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGAT 2711

Qy        361 ATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2712 ATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACG 2771

Qy        421 TCATTCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2772 TCATTCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATT 2831

Qy        481 TCTGGAATGATATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2832 TCTGGAATGATATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCG 2891

Qy        541 CAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2892 CAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGT 2951

Qy        601 GGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2952 GGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGA 3011

Qy        661 TATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3012 TATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCC 3071

Qy        721 AGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3072 AGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTC 3131

Qy        781 ATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3132 ATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGAT 3191

Qy        841 AGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3192 AGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCT 3251

Qy        901 AAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAATGCGGAAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3252 AAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAATGCGGAAA 3311

Qy        961 TAA 963
              |||
Db       3312 TAA 3314
```

RESULT 2
AAH93307/c
ID   AAH93307 standard; DNA; 344 BP.
XX
AC   AAH93307;
XX
DT   04-OCT-2001 (first entry)
XX
DE   Plasmodium falciparum MAL3PC polynucleotide SEQ ID NO 29.
XX
KW   Human; antisense-therapy; gene-therapy; diagnostic; forensic;

KW gene mapping; ds.
XX
OS Plasmodium falciparum.
XX
PN WO200152616-A2.
XX
PD 26-JUL-2001.
XX
PF 22-DEC-2000; 2000WO-US035190.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-451890/48.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.
XX
PS Example 4; Page 113; 135pp; English.
XX
CC The invention relates to an isolated human polynucleotide (AAH75398)
CC encoding a novel polypeptide (AAG64527) useful in antisense-therapy and
CC gene-therapy, in diagnostics, forensics, gene mapping and identification
CC of mutations responsible for genetic disorders and other traits.
CC Polynucleotide sequences with potential homology were also identified
CC (AAH93283-AAH93356)
XX
SQ Sequence 344 BP; 167 A; 16 C; 15 G; 146 T; 0 U; 0 Other;

Query Match 5.8%; Score 55.4; DB 4; Length 344;
Best Local Similarity 56.2%; Pred. No. 0.037;
Matches 104; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```
Qy       759 ATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAA 818
Db       276 ATATATATTATATTTTTATGTATTGGGATTTAATAATAAAAAAAAAGCATATTAAATAAA 217

Qy       819 TAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGCT 878
Db       216 TACTTGTATAATATATATATATATTTTATTCATACATATAATTATATATATATATATATA 157

Qy       879 TAGGTTTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTAT 938
Db       156 TATATATATATATCAACATGGTTTATTTTTTTTTTTTATATTATTATGCGTGAAAAGAT 97

Qy       939 AAAGA 943
Db        96 AAATA 92
```

RESULT 3
ACL35887/c
ID ACL35887 standard; cDNA; 2000 BP.
XX
AC ACL35887;
XX
DT 02-JUN-2005 (first entry)
XX
DE Rice stress-regulated promoter SEQ ID NO:14450.
XX
KW ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW agriculture.
XX
OS Oryza sativa.
XX
PN WO2003008540-A2.
XX
PD 30-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US019668.
XX
PR 22-JUN-2001; 2001US-0300112P.
PR 24-AUG-2001; 2001US-0314662P.
PR 26-SEP-2001; 2001US-0325277P.
PR 21-NOV-2001; 2001US-0332132P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
PI Moughamer T, Provart N, Ricke D, Zhu T;
XX
DR WPI; 2003-248011/24.
XX
PT New stress-responsive nucleic acid, useful for altering the
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT stress, salt stress or osmotic stress.
XX
PS Claim 48; SEQ ID NO 14450; 89pp; English.
XX
CC The invention relates to novel abiotic stress responsive polynucleotides
CC and polypeptides. Also disclosed are vectors, expression cassettes, host
CC cells, and plants containing such polynucleotides. Also disclosed are
CC methods for using the polynucleotides and polypeptides to alter the
CC responsiveness of a plant to abiotic stress. The invention is useful in
CC agriculture. The nucleic acid is useful for determining whether a test
CC plant has been exposed to an abiotic stress condition. It is also useful
CC for selecting an agent that alters abiotic stress regulated
CC polynucleotide expression in a plant cell, and to identify a homolog or
CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC molecule and the polypeptide encoded by it are useful in altering the
CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC stress, osmotic stress or any of their combinations. The present sequence
CC is used in the exemplification of the invention
XX
SQ Sequence 2000 BP; 410 A; 329 C; 253 G; 416 T; 0 U; 592 Other;

Query Match 5.7%; Score 54.8; DB 11; Length 2000;
Best Local Similarity 16.9%; Pred. No. 0.063;
Matches 126; Conservative 298; Mismatches 306; Indels 16; Gaps 5;

```
Qy       229 GATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGC 288
Db       845 GATSTWWAWATRTARCWAWKWRCTGMGTGKTKATKTAWTATAWRKAWMAMARTGCWRMKK 786

Qy       289 GATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGT--TTTTTCGATATTAAAGTCCC 346
Db       785 GWWYAWWWWWTWYWRGKATWAYMMWKYMRMTSRMMYRWARKTTTCYKGWATKATMMAKWS 726

Qy       347 GGCCTGCATTAGATATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATG 406
Db       725 RSKSMSMYGKMSGWYYSSTKWMWGATRMWGMKWGKGCRRCGGYMKWSMKSKWTCASAWGM 666

Qy       407 CTGATGTGAGGACGTCATTCTACACAAATGATGTAGATT-----TTATTTCAGACGTGAA 461
Db       665 TKYKYSYRCRGCGGTWTTRWWMRAWMCMAKWTKTGSYYWCRGYSKYKTTWYWRAWWWSCW 606

Qy       462 AGTTATGTTCACGTTTATTTCTGGAATGATATGTAAGAAAACTGATGCAATTGTCAAAGC 521
Db       605 KKTGTGRTKYMWGTSTYGTTKMYSAYWKRMWSKKRWGRTKYATGWKYWGYTGWYCAYWSA 546

Qy       522 CGTTGGTATTTTCAGTCCGCAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATT 581
Db       545 AMCWKMRKGTKTRTGRKSRGCWRTKCKTYWRMYRWMGYRWRTGTSRGRGGMAATKMSTKY 486

Qy       582 GCCATTACTTAAACAGGGTGGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGA 641
Db       485 RWKRWWRM--AWATKSWKWRGMMWTKWCTKYAYKRWYMRAKAYGKKTTWGCSTYGKCTTW 428

Qy       642 GCCAGATAATTCAGGTGGATATCA-TTTATATAAGGTTTTTTCTAATAATCTTGCGACAA 700
Db       427 AYTRASYGAKWCTRTGKTTKGTCRGMKYRMTTAAGTKAKYYKAKRKTTTKTMWKMRYGSK 368
```

```
Qy        701 TCTTTGATGTTTTTTATCCCAGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCAT 760
              : ::| | | ::: : : :  ::     :: :::::::: :  ::|  : :|:|  ::
Db        367 WRWKTTAGYTRAKWKTYKYTRTTSWWRSKWKKTWWRKKWWRTKGTYKCTGKWRTMAMWWK 308

Qy        761 GTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATA 820
              :|  :|  ::  ::  ::  ::   ::|   |::|  :: : : :|: |    :| : :
Db        307 KTARRTSWWKTKKYTTMKGKRWKCAMRAGTKAKWTCTRWWYYMYWAKRTRTGMMCKRSKG 248

Qy        821 ATTATTTA------AGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATG 874
              :::: | |      | |:|| :  : |   :  : |   :|: | ::::|:::    ::
Db        247 MWWWCTGAAYCACWAYAKATGWYCTKYGRCAMACWTTGTYKTWSARWWYWTVKWTGAYWS 188

Qy        875 GGCTTAGGTTTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATA 934
              :  :|   :::: ::  :| :|  |::::   ::|   : :|:: ::     | ||:| |
Db        187 KCAWTSWYWYWKMWWRMAWAGRTGATWRWMAATYRCACRKCWTKWMWMGGTGAYAAWAAA 128

Qy        935 TTATAAAGACGGTTTTAATGCGGAAA 960
              ||:::: ::   | |:: : :: :::
Db        127 TTRKRMKRMATWTMTYWKKMYRMWWW 102
```

RESULT 4
ABN80057
ID ABN80057 standard; DNA; 10652 BP.
XX
AC ABN80057;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human chemically modified disease associated gene SEQ ID NO 74.
XX
KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cytostatic; anticonvulsant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200927-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007536.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130908/17.
XX
PT Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a sequence
PT of a segment of chemically pretreated DNA of genes associated with
PT development.
XX
PS Claim 1; SEQ ID NO 74; 27pp; English.
XX
CC The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in the
CC specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with
CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Currarino syndrome, diseases related with the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (II) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (II) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
CC patent did not form part of the printed specification but is based on
CC sequence information supplied to Derwent by the European Patent Office
XX
SQ Sequence 10652 BP; 2166 A; 601 C; 3168 G; 4717 T; 0 U; 0 Other;

Query Match 5.5%; Score 53; DB 6; Length 10652;
Best Local Similarity 46.6%; Pred. No. 0.19;
Matches 170; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

```
Qy        590 TTAAACAGGGTGGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATA 649
              | |||     || |   |||      |||     |  || | | |||||||        ||||
Db       5600 TAAAATTAAGTCGTTCGTTTTTTTTTAATTTTTTTATTTTGATTGAGGAGTTTTTAGATC 5659

Qy        650 ATTCAGGTGGATATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATG 709
              |  |  | | | |  ||  | |||  ||||||  | ||||  |||     |   | | |
Db       5660 GGTTATTTCGTTTTTTTTGTTTTAAAATTTTTTAAATTAATTTTGTCGGAGAGTAGGGTT 5719

Qy        710 TTTTTTATCCCAGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCT 769
              | ||| |  | |||      | |||    ||| ||  | |  | ||   |||| || |
Db       5720 TATTTGAAACGAGAATTAGGAATTTAACGGTAAGGGTGGGAGTTTTAGTTGGTTTTTTTT 5779

Qy        770 TACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAA 829
              |  ||   ||  ||      |  ||||          | || |  |   |||  ||| ||
Db       5780 TTTTTTTTTTTTTATTTTTTTTAGATTTTTTTTTTAGTTGTTGTTAGAATGTTTAATTGG 5839

Qy        830 GAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCT 889
              || ||    |  |   ||| |   |||  || |       | |      |  |||||| |
Db       5840 AAGTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTCGGGTTTTTTTTTATTTGTTTTTGT 5899

Qy        890 ATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTT 949
              | | |||   | |   ||   | | ||  |    |   |||  | ||||      |||
Db       5900 TTATTTATTAAATATTCGTTGTTATTTTAGGTTCGTTTGTTTATTATTTAAAATGTATTT 5959

Qy        950 TAATG 954
              | |||
Db       5960 TTATG 5964
```

RESULT 5
AAA53723
ID AAA53723 standard; DNA; 906 BP.
XX
AC AAA53723;
XX
DT 15-SEP-2003 (revised)
DT 22-DEC-2000 (first entry)
XX
DE Campylobacter jejuni OH4384 Beta-1,3-galactosyltransferase.
XX
KW Biosynthetic locus; biosynthesis; lipid A biosynthesis;
KW acetyltransferase; glycosyltransferase; Beta-1,4-GalNAc transferase;
KW Beta-1,3-galactosyltransferase; alpha-2,3-sialyltransferase;
KW sialic acid synthase; CMP-sialic acid synthetase; mimic; antibody;
KW immunity; immunogen; ganglioside; ds.
XX
OS Campylobacter jejuni; OH4384.
XX
FH Key Location/Qualifiers
FT CDS 1. .906
FT /*tag= a
FT /product= "Beta-1,4-galactosyltransferase"
XX
PN WO200046379-A1.
XX
PD 10-AUG-2000.
XX
PF 01-FEB-2000; 2000WO-CA000086.

XX
PR 01-FEB-1999; 99US-0118213P.
PR 31-JAN-2000; 2000US-00495406.
XX
PA (CANA) NAT RES COUNCIL CANADA.
XX
PI Gilbert M, Wakarchuk WW;
XX
DR WPI; 2000-524418/47.
DR P-PSDB; AAY97206.
XX
PT Novel glycosyltransferase polypeptides and polynucleotides useful for
PT biosynthesis of ganglioside and ganglioside mimics, as diagnostic
PT reagents and as immunogen for producing antibodies.
XX
PS Claim 8; Page 101-102; 120pp; English.
XX
CC A reaction mixture for the synthesis of a sialylated oligosaccharide is
CC useful for synthesising sialylated oligosaccharide such as ganglioside,
CC lysoganglioside or their mimics. Glycosyltransferases are useful for
CC chemo-enzymatic synthesis of oligosaccharides, including gangliosides and
CC other oligosaccharides that have biological activity. The enzymes and
CC nucleic acids that encode them are useful for studies of the pathogenesis
CC mechanisms of organisms that synthesize ganglioside mimics, such as C.
CC jejuni and the nucleic acids are used as probes to study expression of
CC genes involved in ganglioside mimetic synthesis. Antibodies raised
CC against the glycosyltransferases are also useful for analyzing the
CC expression patterns of these genes involved in pathogenesis. The nucleic
CC acids are also useful for designing antisense oligonucleotides for
CC inhibiting expression of the Campylobacter enzymes that are involved in
CC the biosynthesis of ganglioside mimics that can mask the pathogens from
CC the host's immune system. The oligosaccharides are useful as diagnosing
CC reagents or as therapeutics and as immunogens for producing antibodies.
CC Bacterial glycosyltransferase can be used to catalyse the formation of
CC oligosaccharides that are identical to the corresponding mammalian
CC structures and are easier and less expensive to produce in large
CC quantity, compared to the mammalian glycosyltransferase. The bacterial
CC origin of the enzymes facilitates expression of large quantities of the
CC enzymes using relatively inexpensive prokaryotic expression systems.
CC (Updated on 15-SEP-2003 to standardise OS field)
XX
SQ Sequence 906 BP; 377 A; 79 C; 137 G; 313 T; 0 U; 0 Other;

Query Match 5.3%; Score 50.6; DB 3; Length 906;
Best Local Similarity 49.1%; Pred. No. 0.43;
Matches 134; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
Qy      105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
            | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db       87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy      165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
            |||      ||   | |||   | ||  |  |        |  |||||| ||||||||||
Db      147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy      225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |      |         |  || || |  |  || ||||       ||| |   | |  |
Db      207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy      285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
             |    |||||||    | |    ||||| |  |    ||  | |    || ||| |
Db      267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy      345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
             | |  || | | ||| | |||| | | ||| |
Db      327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359
```

RESULT 6
ABT13678
ID ABT13678 standard; DNA; 906 BP.
XX
AC ABT13678;

XX
DT 07-FEB-2003 (first entry)
XX
DE Campylobacter jejuni Beta-1,3-galactosyltransferase coding sequence #1.
XX
KW Gene; ds; enzyme; gene therapy; acyltransferase; glycosyltransferase;
KW GalNAc transferase; N-Acetylgalactosamine transferase;
KW galactosyltransferase; sialyltransferase; sialic acid synthase;
KW cytidine 5'-monophosphate sialic acid synthetase;
KW CMP sialic acid synthetase; acetyltransferase; ganglioside synthesis;
KW ganglioside mimetics; inflammation; tumour metastasis.
XX
OS Campylobacter jejuni.
XX
PN WO200274942-A2.
XX
PD 26-SEP-2002.
XX
PF 22-FEB-2002; 2002WO-CA000229.
XX
PR 21-MAR-2001; 2001US-00816028.
XX
PA (CANA) NAT RES COUNCIL CANADA.
XX
PI Gilbert M, Wakarchuk WW;
XX
DR WPI; 2003-040554/03.
DR P-PSDB; ABJ18491.
XX
PT New glycosyltransferases from Campylobacter, useful for synthesizing
PT gangliosides and ganglioside mimetics, and in studying the pathogenesis
PT mechanisms of organisms that synthesize ganglioside mimetics.
XX
PS Claim 8; Page 104; 107pp; English.
XX
CC The invention comprises the amino acid and coding sequences of
CC Campylobacter jejuni proteins. The C. jejuni proteins of the invention
CC may be either an: acyltransferase; glycosyltransferase; GalNAc (N-
CC Acetylgalactosamine) transferase; galactosyltransferase;
CC sialyltransferase; sialic acid synthase; cytidine 5'-monophosphate (CMP)
CC sialic acid synthetase; acetyltransferase. The C. jejuni DNA and protein
CC sequences of the invention are useful for ganglioside synthesis, studying
CC ganglioside mimetics, and for designing oligonucleotides to inhibit
CC expression of Campylobacter enzymes involved in the biosynthesis of
CC ganglioside mimetics that can mask the pathogen's from the host's immune
CC system. The C. jejuni oligosaccharides of the invention may be used as
CC diagnostic reagents (e.g. to locate areas of inflammation or tumour
CC metastasis). The present DNA sequence represents a Campylobacter jejuni
CC gene of the invention
XX
SQ Sequence 906 BP; 377 A; 79 C; 137 G; 313 T; 0 U; 0 Other;

Query Match 5.3%; Score 50.6; DB 8; Length 906;
Best Local Similarity 49.1%; Pred. No. 0.43;
Matches 134; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
Qy      105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
            | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db       87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy      165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
            |||      ||   | |||   | ||  |  |        |  |||||| ||||||||||
Db      147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy      225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |      |         |  || || |  |  || ||||       ||| |   | |  |
Db      207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy      285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
             |    |||||||    | |    ||||| |  |    ||  | |    || ||| |
Db      267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy      345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
```

```
                    | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 7
AAA53720/c
ID   AAA53720 standard; DNA; 11474 BP.
XX
AC   AAA53720;
XX
DT   15-SEP-2003  (revised)
DT   22-DEC-2000  (first entry)
XX
DE   LPS core biosynthesis locus.
XX
KW   Biosynthetic locus; biosynthesis; lipid A biosynthesis;
KW   acetyltransferase; glycosyltransferase; Beta-1,4-GalNAc transferase;
KW   Beta-1,3-galactosyltransferase; alpha-2,3-sialyltransferase;
KW   sialic acid synthase; CMP-sialic acid synthetase; mimic; antibody;
KW   immunity; immunogen; ganglioside; ds.
XX
OS   Campylobacter jejuni; OH4384.
XX
FH   Key             Location/Qualifiers
FT   CDS             350. .1237
FT                   /*tag= a
FT                   /product= "acyltransferase"
FT                   /note= "Open reading frame 2b"
FT   CDS             1234. .2490
FT                   /*tag= b
FT                   /product= "glycosyl transferase"
FT                   /note= "Open reading frame 3a"
FT   CDS             2786. .3955
FT                   /*tag= c
FT                   /product= "Glycosyl transferase"
FT                   /note= "Open reading frame 4b"
FT   CDS             4025. .5068
FT                   /*tag= d
FT                   /product= "Beta 1,4-GalNAc transferase"
FT                   /note= "Open reading frame 5a"
FT   CDS             5054. .5959
FT                   /*tag= e
FT                   /product= "Beta 1,2- galactosyltransferase"
FT                   /note= "Open reading frame 6a"
FT   CDS             6048. .6923
FT                   /*tag= f
FT                   /product= "CstII sialyltransferase"
FT                   /note= "Open reading frame 7a"
FT   CDS             6924. .7964
FT                   /*tag= g
FT                   /product= "Sialic acid synthase"
FT                   /note= "Open reading frame 8a"
FT   CDS             8020. .9079
FT                   /*tag= h
FT                   /product= "Sialic acid biosynthetic enzyme"
FT                   /note= "Open reading frame 9a"
FT   CDS             9076. .9741
FT                   /*tag= i
FT                   /product= "CMP-sialic acid synthetase"
FT                   /note= "Open reading frame 10a"
FT   CDS             complement(10554. .11366)
FT                   /*tag= j
FT                   /product= "Lipooligosaccharide biosynthetic enzyme"
FT                   /note= "Open reading frame 12a"
XX
PN   WO200046379-A1.
XX
PD   10-AUG-2000.
XX
PF   01-FEB-2000; 2000WO-CA000086.
XX
PR   01-FEB-1999;   99US-0118213P.
PR   31-JAN-2000; 2000US-00495406.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Gilbert M,  Wakarchuk WW;
XX
DR   WPI; 2000-524418/47.
DR   P-PSDB; AAY97200, AAY97201, AAY97202, AAY97203, AAY97204, AAY97205,
DR   AAY97206.
XX
PT   Novel glycosyltransferase polypeptides and polynucleotides useful for
PT   biosynthesis of ganglioside and ganglioside mimics, as diagnostic
PT   reagents and as immunogen for producing antibodies.
XX
PS   Claim 1; Page 86-90; 120pp; English.
XX
CC   A reaction mixture for the synthesis of a sialylated oligosaccharide is
CC   useful for synthesising sialylated oligosaccharide such as ganglioside,
CC   lysoganglioside or their mimics. Glycosyltransferases are useful for
CC   chemo-enzymatic synthesis of oligosaccharides, including gangliosides and
CC   other oligosaccharides that have biological activity. The enzymes and
CC   nucleic acids that encode them are useful for studies of the pathogenesis
CC   mechanisms of organisms that synthesize ganglioside mimics, such as C.
CC   jejuni and the nucleic acids are used as probes to study expression of
CC   genes involved in ganglioside mimetic synthesis. Antibodies raised
CC   against the glycosyltransferases are also useful for analyzing the
CC   expression patterns of these genes involved in pathogenesis. The nucleic
CC   acids are also useful for designing antisense oligonucleotides for
CC   inhibiting expression of the Campylobacter enzymes that are involved in
CC   the biosynthesis of ganglioside mimics that can mask the pathogens from
CC   the host's immune system. The oligosaccharides are useful as diagnosing
CC   reagents or as therapeutics and as immunogens for producing antibodies.
CC   Bacterial glycosyltransferase can be used to catalyse the formation of
CC   oligosaccharides that are identical to the corresponding mammalian
CC   structures and are easier and less expensive to produce in large
CC   quantity, compared to the mammalian glycosyltransferase. The bacterial
CC   origin of the enzymes facilitates expression of large quantities of the
CC   enzymes using relatively inexpensive prokaryotic expression systems.
CC   (Updated on 15-SEP-2003 to standardise OS field)
XX
SQ   Sequence 11474 BP; 4506 A; 1495 C; 1613 G; 3860 T; 0 U; 0 Other;

  Query Match             5.3%;  Score 50.6;  DB 3;  Length 11474;
  Best Local Similarity   49.1%;  Pred. No. 0.61;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   ||||  ||  |  ||  ||| | |    ||  |||||
Db       5873 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 5814

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db       5813 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 5754

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |    |          |  || || |  |  || ||||       ||| |   | | |
Db       5753 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 5694

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |  |||||||    | |    ||||| |  |    ||  | |    || ||| |
Db       5693 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 5634

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
                | |  || | | ||| | |||| | | ||| |
Db       5633 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 5601


RESULT 8
ABT13665/c
ID   ABT13665 standard; DNA; 11474 BP.
XX
AC   ABT13665;
XX
DT   07-FEB-2003  (first entry)
```

XX
DE Campylobacter jejuni genomic lipooligosaccharide biosynthesis locus.
XX
KW Gene; ds; enzyme; gene therapy; acyltransferase; glycosyltransferase;
KW GalNAc transferase; N-Acetylgalactosamine transferase;
KW galactosyltransferase; sialyltransferase; sialic acid synthase;
KW cytidine 5'-monophosphate sialic acid synthetase;
KW CMP sialic acid synthetase; acetyltransferase; ganglioside synthesis;
KW ganglioside mimetics; inflammation; tumour metastasis.
XX
OS Campylobacter jejuni.
XX
PN WO200274942-A2.
XX
PD 26-SEP-2002.
XX
PF 22-FEB-2002; 2002WO-CA000229.
XX
PR 21-MAR-2001; 2001US-00816028.
XX
PA (CANA) NAT RES COUNCIL CANADA.
XX
PI Gilbert M, Wakarchuk WW;
XX
DR WPI; 2003-040554/03.
XX
PT New glycosyltransferases from Campylobacter, useful for synthesizing
PT gangliosides and ganglioside mimetics, and in studying the pathogenesis
PT mechanisms of organisms that synthesize ganglioside mimetics.
XX
PS Claim 1; Page 89-95; 107pp; English.
XX
CC The invention comprises the amino acid and coding sequences of
CC Campylobacter jejuni proteins. The C. jejuni proteins of the invention
CC may be either an: acyltransferase; glycosyltransferase; GalNAc (N-
CC Acetylgalactosamine) transferase; galactosyltransferase;
CC sialyltransferase; sialic acid synthase; cytidine 5'-monophosphate (CMP)
CC sialic acid synthetase; acetyltransferase. The C. jejuni DNA and protein
CC sequences of the invention are useful for ganglioside synthesis, studying
CC ganglioside mimetics, and for designing oligonucleotides to inhibit
CC expression of Campylobacter enzymes involved in the biosynthesis of
CC ganglioside mimetics that can mask the pathogen's from the host's immune
CC system. The C. jejuni oligosaccharides of the invention may be used as
CC diagnostic reagents (e.g. to locate areas of inflammation or tumour
CC metastasis). The present DNA sequence represents a Campylobacter jejuni
CC gene of the invention
XX
SQ Sequence 11474 BP; 4506 A; 1495 C; 1613 G; 3860 T; 0 U; 0 Other;

Query Match 5.3%; Score 50.6; DB 8; Length 11474;
Best Local Similarity 49.1%; Pred. No. 0.61;
Matches 134; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
Qy       105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
Db      5873 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 5814

Qy       165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
Db      5813 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 5754

Qy       225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
Db      5753 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 5694

Qy       285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
Db      5693 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 5634

Qy       345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
Db      5633 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 5601
```

RESULT 9
ADZ70897
ID ADZ70897 standard; DNA; 1524 BP.
XX
AC ADZ70897;
XX
DT 14-JUL-2005 (first entry)
XX
DE Human matrix attachment region DNA SEQ ID NO 21.
XX
KW ds; matrix attachment region; MAR; protein production.
XX
OS Homo sapiens.
XX
PN WO2005040377-A2.
XX
PD 06-MAY-2005.
XX
PF 22-OCT-2004; 2004WO-EP011974.
XX
PR 24-OCT-2003; 2003US-0513574P.
PR 06-FEB-2004; 2004EP-00002722.
XX
PA (SELE-) SELEXIS SA.
XX
PI Mermod N, Girod PA, Bucher P, Nguyen D, Calabrese D, Saugy D;
PI Puttini S;
XX
DR WPI; 2005-333507/34.
XX
PT New purified and isolated DNA sequence having protein production
PT increasing activity comprises a bent DNA element and a binding site for a
PT DNA binding protein, useful for increasing protein production activity in
PT eukaryotic host cell.
XX
PS Claim 4; SEQ ID NO 21; 282pp; English.
XX
CC The invention relates to a purified and isolated DNA sequence having
CC protein production increasing activity comprising at least one bent DNA
CC element, and at least one binding site for a DNA binding protein. The
CC purified and isolated DNA sequence comprising a first and a second
CC isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC MAR nucleotide sequence selected from a purified and isolated DNA
CC sequence above, a purified and isolated MAR DNA above, a purified and
CC isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC sequence complementary to it, its molecular chimera, or its combinations
CC and variants, is useful for increasing protein production activity in a
CC eukaryotic host cell. The present sequence represents a human matrix
CC attachment region DNA.
XX
SQ Sequence 1524 BP; 943 A; 14 C; 15 G; 552 T; 0 U; 0 Other;

Query Match 5.2%; Score 49.6; DB 14; Length 1524;
Best Local Similarity 57.4%; Pred. No. 0.75;
Matches 108; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

```
Qy       757 GCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACA 816
Db       501 GAATATATATGAATATATAAATATATAAAAAATATATATAAATATATAAATATATATAAA 560

Qy       817 AATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGG 876
Db       561 TATAAATATATAAAAATATATAAAAATATATATAAATATATAAATATATATAAATATATA 620

Qy       877 CTTAGGTTTTTCTATCTAT-ATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATAT 935
Db       621 AATATATATAAATATATATAAATATATAAATATATATAAATATATATAAATATATAAATA 680

Qy       936 TATAAAGA 943
Db       681 TATAAATA 688
```

```
RESULT 10
ADZ70902
ID   ADZ70902 standard; DNA; 4660 BP.
XX
AC   ADZ70902;
XX
DT   14-JUL-2005  (first entry)
XX
DE   Human matrix attachment region DNA SEQ ID NO 26.
XX
KW   ds; matrix attachment region; MAR; protein production.
XX
OS   Homo sapiens.
XX
PN   WO2005040377-A2.
XX
PD   06-MAY-2005.
XX
PF   22-OCT-2004; 2004WO-EP011974.
XX
PR   24-OCT-2003; 2003US-0513574P.
PR   06-FEB-2004; 2004EP-00002722.
XX
PA   (SELE-) SELEXIS SA.
XX
PI   Mermod N,  Girod PA,  Bucher P,  Nguyen D,  Calabrese D,  Saugy D;
PI   Puttini S;
XX
DR   WPI; 2005-333507/34.
XX
PT   New purified and isolated DNA sequence having protein production
PT   increasing activity comprises a bent DNA element and a binding site for a
PT   DNA binding protein, useful for increasing protein production activity in
PT   eukaryotic host cell.
XX
PS   Claim 4; SEQ ID NO 26; 282pp; English.
XX
CC   The invention relates to a purified and isolated DNA sequence having
CC   protein production increasing activity comprising at least one bent DNA
CC   element, and at least one binding site for a DNA binding protein. The
CC   purified and isolated DNA sequence comprising a first and a second
CC   isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC   MAR nucleotide sequence selected from a purified and isolated DNA
CC   sequence above, a purified and isolated MAR DNA above, a purified and
CC   isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC   sequence complementary to it, its molecular chimera, or its combinations
CC   and variants, is useful for increasing protein production activity in a
CC   eukaryotic host cell. The present sequence represents a human matrix
CC   attachment region DNA.
XX
SQ   Sequence 4660 BP; 1770 A; 774 C; 775 G; 1341 T; 0 U; 0 Other;

  Query Match             5.2%;  Score 49.6;  DB 14;  Length 4660;
  Best Local Similarity   57.4%;  Pred. No. 0.87;
  Matches  108;  Conservative    0;  Mismatches   79;  Indels    1;  Gaps    1;

Qy        757 GCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACA 816
              | || | |||   ||  ||| |  |||    ||  | |||||   | |||| |   || |
Db       1510 GAATATATATGAATATATAAATATATAAAAAATATATATAAATATATAAATATATATAAA 1569

Qy        817 AATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGG 876
               |||| ||| ||| |    |   | | |  || || | || ||| ||||||||||||
Db       1570 TATAAATATATAAAAATATATAAAAATATATATAAATATATAAATATATATAAATATATA 1629

Qy        877 CTTAGGTTTTTCTATCTAT-ATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATAT 935
                ||  | |   ||| ||| |   || || | |   |  || || | ||||| |  |
Db       1630 AATATATATAAATATATATAAATATATAAATATATATAAATATATATAAATATATAAATA 1689

Qy        936 TATAAAGA 943
              |||||| |
Db       1690 TATAAATA 1697
```

```
RESULT 11
AAS46623
ID   AAS46623 standard; DNA; 5654 BP.
XX
AC   AAS46623;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #345.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 345; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5654 BP; 1036 A; 255 C; 1591 G; 2772 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 48.4;  DB 4;  Length 5654;
  Best Local Similarity   47.7%;  Pred. No. 1.6;
  Matches  142;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy        666 TTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGA 725
              |||  | | |  ||||||  | |||       || ||| ||| || |||    |     |
Db        383 TTTTTTTTAATTTTTTTAAATTAAGAAGAGGATAATTTTTAATTTTTAGGTTTTTTAAAT 442
```

```
Qy        726 GCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCATAGG 785
              |     | ||   ||     ||    |  |||||  ||  |||| ||  || |
Db        443 TTTTTTTTCGTTTAATTTTTTTAATTATTGTTAATTTTATGGTTTTTTAAATTTTTAAG 502

Qy        786 CGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGATAGTGC 845
              ||| |  |||    | |||||||  |   ||  |||||||||   ||     | |
Db        503 GTATGTATTTAAGGTTATAAATTTTTTTTTTAAATATTATTTAAATTATATTTTTTATAT 562

Qy        846 ATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCTAAAAC 905
              || ||| ||   |||  || | ||| |    |||  ||  | | | | ||   | |
Db        563 TTTGATATATGGTATTTAATTTTAATTTTTATTTATTTTAGTATTTTTTTAATTTTTATT 622

Qy        906 TGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAATGCGGAAATAA 963
              |||  ||      |  |  | | |  ||| |  | |||| |   |  |||| 
Db        623 ATGATTTTTTGTTTATTTTAGGAGTTTTGTAGTAATGTGGTTTTTGTTATTGTTATAA 680
```

RESULT 12
ABL33874
ID ABL33874 standard; DNA; 5654 BP.
XX
AC ABL33874;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1847.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1847; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 5654 BP; 1036 A; 255 C; 1591 G; 2772 T; 0 U; 0 Other;

Query Match 5.0%; Score 48.4; DB 6; Length 5654;
Best Local Similarity 47.7%; Pred. No. 1.6;
Matches 142; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

```
Qy        666 TTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGA 725
              ||| | | |  ||||||  | |||       || ||| ||| || |||     |      |
Db        383 TTTTTTTTAATTTTTTTAAATTAAGAAGAGGATAATTTTTAATTTTTAGGTTTTTTAAAT 442

Qy        726 GCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCATAGG 785
              |     | ||   ||     ||    |  |||||  ||  |||| ||  || |
Db        443 TTTTTTTTCGTTTAATTTTTTTAATTATTGTTAATTTTATGGTTTTTTAAATTTTTAAG 502

Qy        786 CGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGATAGTGC 845
              ||| |  |||    | |||||||  |   ||  |||||||||   ||     | |
Db        503 GTATGTATTTAAGGTTATAAATTTTTTTTTTAAATATTATTTAAATTATATTTTTTATAT 562

Qy        846 ATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCTAAAAC 905
              || ||| ||   |||  || | ||| |    |||  ||  | | | | ||   | |
Db        563 TTTGATATATGGTATTTAATTTTAATTTTTATTTATTTTAGTATTTTTTTAATTTTTATT 622

Qy        906 TGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAATGCGGAAATAA 963
              |||  ||      |  |  | | |  ||| |  | |||| |   |  |||| 
Db        623 ATGATTTTTTGTTTATTTTAGGAGTTTTGTAGTAATGTGGTTTTTGTTATTGTTATAA 680
```

RESULT 13
ABQ67159
ID ABQ67159 standard; DNA; 6775 BP.
XX
AC ABQ67159;
XX
DT 28-AUG-2002 (first entry)
XX
DE Human angiogenesis associated polynucleotide SEQ ID NO 189.
XX
KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.
XX
OS Homo sapiens.
XX
PN WO200246454-A2.
XX
PD 13-JUN-2002.
XX
PF 06-DEC-2001; 2001WO-EP014320.
XX
PR 06-DEC-2000; 2000DE-01061338.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Schacht O;
XX
DR WPI; 2002-500450/53.
XX
PT New nucleic acid fragments from chemically treated angiogenesis-
PT associated genes, useful for determining methylation status, e.g. in
PT diagnosis or treatment of cancer.
XX
PS Claim 1; SEQ ID NO 189; 41pp + Sequence Listing; German.
XX
CC The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

```
SQ   Sequence 6775 BP; 1982 A; 49 C; 1220 G; 3524 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 48;  DB 6;  Length 6775;
  Best Local Similarity   56.2%;  Pred. No. 2;
  Matches   90;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy        781 ATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGAT 840
              |||   || | || |  ||    |||   |  | ||| |||||||| | ||| ||  ||
Db       2535 ATAATAGAGGGAGTTTTAATTTTAAAAAATTAAAAAAAAAATTATTGAGGAGTTTAAGAG 2594

Qy        841 AGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCT 900
              |  ||||||   ||||  | |||  | ||| |     |    |  |     || |   |
Db       2595 TTTTTATTTATGAGTTTATATTTATTAATATTTAATGGTTAAGTGATTATGTTAAAGTTT 2654

Qy        901 AAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAA 940
              |||| | |   ||||||||||||||  | |||||  |||
Db       2655 AAAAATATTTATTTATTTAGAAAAATTTATTATATGTTAA 2694


RESULT 14
ACL37108
ID   ACL37108 standard; cDNA; 2000 BP.
XX
AC   ACL37108;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Rice stress-regulated promoter SEQ ID NO:15671.
XX
KW   ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW   agriculture.
XX
OS   Oryza sativa.
XX
PN   WO2003008540-A2.
XX
PD   30-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US019668.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   24-AUG-2001; 2001US-0314662P.
PR   26-SEP-2001; 2001US-0325277P.
PR   21-NOV-2001; 2001US-0332132P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Kreps J,  Briggs SP,  Cooper B,  Glazebrook J,  Goff SA,  Katagiri F;
PI   Moughamer T,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-248011/24.
XX
PT   New stress-responsive nucleic acid, useful for altering the
PT   responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT   stress, salt stress or osmotic stress.
XX
PS   Claim 48; SEQ ID NO 15671; 89pp; English.
XX
CC   The invention relates to novel abiotic stress responsive polynucleotides
CC   and polypeptides. Also disclosed are vectors, expression cassettes, host
CC   cells, and plants containing such polynucleotides. Also disclosed are
CC   methods for using the polynucleotides and polypeptides to alter the
CC   responsiveness of a plant to abiotic stress. The invention is useful in
CC   agriculture. The nucleic acid is useful for determining whether a test
CC   plant has been exposed to an abiotic stress condition. It is also useful
CC   for selecting an agent that alters abiotic stress regulated
CC   polynucleotide expression in a plant cell, and to identify a homolog or
CC   ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC   molecule and the polypeptide encoded by it are useful in altering the
CC   responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC   stress, osmotic stress or any of their combinations. The present sequence
CC   is used in the exemplification of the invention
XX

SQ   Sequence 2000 BP; 510 A; 350 C; 268 G; 518 T; 0 U; 354 Other;

  Query Match             4.9%;  Score 47.4;  DB 11;  Length 2000;
  Best Local Similarity   15.1%;  Pred. No. 2.3;
  Matches   76;  Conservative  217;  Mismatches  204;  Indels    7;  Gaps    3;

Qy        433 AATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATTTCTGGAATGATA 492
              || :::  :   :: ::: :|:: :    ::::  :|::   :|::|  ::: ::|::
Db          3 AAARRYWRRCSSWKMMWWMMARRRSAMYKWSYYYSRTKYYKWYYYRTCSYKRWWMTSMMK 62

Qy        493 TGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCGCAAACTACTGGA 552
              ::  : :: : :  :    |:  ::|  :: ::   :| |: : :: :  ::   : : |
Db         63 YRAYMMRWYWKKMWAMWYAWCMWYAYGYMMWKWWYSRTMTYTRAYYKKWKMWRRKMSRMA 122

Qy        553 AAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGTGGA---GAGTTC 609
              ||:: :::::|: :::|    ::  :|   :::|::|:|: :::  :: ::   :| :::
Db        123 AAWYYWSYWRTSRWKWTGYRKKRWWRARRYMYWTYWCYTWSWMRWYRYWKMMKYRAYKWM 182

Qy        610 GCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGATATCATTTA 669
               :  |::| |  :::  : : ::: | :    :  : : :: :    |  ||:  :   |
Db        183 AYTTTYMTKCCMWWWCTYGYKWKWAAWRASAKRARRAWMMMRWKSMWTWAATWSWWSMMA 242

Qy        670 TATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGAGCAC 729
              |:::  :: : : :   ::: :: |  :| :: : || |  |:::::  :   :: :|
Db        243 TWKRKMRWGWMWAYSSTWMMRWWWYCCTRRAMYKMKAATMTRCTWYMWMAAMTSAWRAWC 302

Qy        730 CGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCATAGGCGAT 789
               |: |   | : | : |:::: :::| :|:  :|  ::  : :::::|    |:: | :
Db        303 AGWATSYATKWYAWRTTYMKYRRRYYACMAWTRWTKMWSMMRTKKWWYTM---ARSWGCK 359

Qy        790 GAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGATAGTGCATTT 849
              ::: |:|::|  : :|  : |::::| :::::: :   : :|: ::  : :  ::|| ::
Db        360 KWMWAKARMATMCMAYTYSTYRYYAWRRRWMRWSWAA-WYSAKYWYAKSYYYMYKCARKW 418

Qy        850 ATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCTAAAACTGTG 909
              :|  |:: || :||:||  :   :    |:|:  :: : : ::: : | : ::: ::  :
Db        419 WTTYAYYAAAAKATRTAAYMGRTWYTMCWTMYTKYMKCMRWWYMMCAMYKKWMGWMAYWT 478

Qy        910 CCTTTATTTAGAAAAATAAAATAT 933
              |:|   |  :  | |  ||||
Db        479 GGTYTCAGGACTMNCACATCATAT 502


RESULT 15
ABK40038/c
ID   ABK40038 standard; DNA; 16258 BP.
XX
AC   ABK40038;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Human chemically pretreated gene sequence #60 strand 2.
XX
KW   Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW   cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW   UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO200202806-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007470.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154757/20.
```

XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 120; 24pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 16258 BP; 4242 A; 330 C; 3801 G; 7885 T; 0 U; 0 Other;

Query Match 4.9%; Score 47; DB 6; Length 16258;
Best Local Similarity 48.4%; Pred. No. 3.6;
Matches 162; Conservative 0; Mismatches 170; Indels 3; Gaps 1;

```
Qy        620 ATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGATATCATTTATATAAGGTTT 679
              || | |  ||||   |     |   | | ||||    |  || | ||||||  |  | |
Db       2690 ATCAAACAATAAAACAACTATAAAAAAAAAATTTTTATCTATCTTATTTATTCACCGCTA 2631

Qy        680 TTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGAGCACCGTGTAAGTA 739
              | ||    |  |||   |||||  ||  |    || | | |  |    |     ||| ||
Db       2630 TATCCCCAACACTTAAAACAATACTTATTCACATTAAACGCCCAATATATTAAATAATTA 2571

Qy        740 AAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAA 799
              ||  | |      |  |  |  |  ||   |||  ||    |         || |||||
Db       2570 AATAACTAAATAAACGAATAAATAAATAAATACACAAAACGAATCAAATAAAAAATAAAC 2511

Qy        800 CCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAA 859
              | ||    ||  || | |||  ||| ||  | |     |   |  || |||| | | ||
Db       2510 CAAACCTATTATATATACATATATATATATAAAAAATAAACTATATATATATAAAATAAA 2451

Qy        860 TTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCTAAAACTGTGCCTT---TAT 916
              |||||||| |||     |         | ||| |||||   |||  | |    |   |||
Db       2450 CTATATATATATAATATATATATAAAATATATATATATAAAAAATATATAAACTATATAT 2391

Qy        917 TTAGAAAAATAAAATATATTATAAAGACGGTTTTA 951
               || | | |||| |||||| ||| | |   || ||
Db       2390 ATATATATATAATATATATAATATATATATTTCTA 2356
```

Search completed: May 11, 2006, 09:42:38
Job time : 442.793 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 09:43:19 ; Search time 3000.39 Seconds
(without alignments)
15016.694 Million cell updates/sec

Title: US-09-423-093B-3_COPY_2352_3314
Perfect score: 963
Sequence: 1 atgagtaaaaaacctcttct.....cggttttaatgcggaaataa 963

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	78.6	8.2	1101	10	CNS0039G	AL063921 Drosophil
c 2	59.2	6.1	987	10	CNS014PQ	AL104456 Drosophil
3	58.6	6.1	1101	10	CNS00EVL	AL069706 Drosophil
c 4	58.6	6.1	1101	10	CNS00EVL	AL069706 Drosophil
5	57.8	6.0	1201	10	CNS010D6	AL098820 Drosophil
c 6	57.6	6.0	1101	10	CNS00EPO	AL069493 Drosophil
7	57	5.9	928	10	CNS00DKY	AL071865 Drosophil
c 8	54.8	5.7	1272	9	CC264939	CC264939 CH261-19L
9	54.6	5.7	1160	1	AJ926878	AJ926878 AJ926878
10	54.4	5.6	994	10	CNS0058N	AL057169 Drosophil
c 11	54.2	5.6	928	10	CNS00DKY	AL071865 Drosophil
12	54	5.6	1101	10	CNS004ZW	AL055440 Drosophil
13	53.8	5.6	751	9	BH250178	BH250178 BOGAF63TF
c 14	53.4	5.5	945	11	CNS04D0K	AL285149 Tetraodon
15	53.2	5.5	706	9	BH955189	BH955189 odi49d09.
16	53.2	5.5	858	10	AG877429	AG877429 Oryza sat
c 17	52.4	5.4	1101	10	CNS001FB	AL060732 Drosophil
18	52.2	5.4	354	3	BI814524	BI814524 PfESToaa2
c 19	52.2	5.4	1076	1	AJ925668	AJ925668 AJ925668
c 20	51.8	5.4	500	1	AU086081	AU086081 AU086081
c 21	51.8	5.4	1101	10	CNS003BD	AL064091 Drosophil
22	51.8	5.4	1101	10	CNS0106X	AL098595 Drosophil
23	51.6	5.4	500	1	AU088184	AU088184 AU088184
24	51.6	5.4	950	9	BH860748	BH860748 ORNL247 P
c 25	51.4	5.3	575	10	CW277980	CW277980 104_752_1
26	51.4	5.3	580	3	BI814403	BI814403 PfESToaa2
27	51.4	5.3	805	9	BZ085876	BZ085876 l1a16h04.
c 28	51	5.3	567	10	CL509924	CL509924 SAIL_821_
c 29	50.8	5.3	1101	10	CNS00FVE	AL071298 Drosophil
30	50.6	5.3	581	8	DN470449	DN470449 USDA-FP_1
31	50.6	5.3	581	8	DR755050	DR755050 USDA-FP_1
c 32	50.6	5.3	1101	10	CNS004ZW	AL055440 Drosophil
c 33	50.6	5.3	1201	10	CNS016E1	AL106627 Drosophil
c 34	50.4	5.2	1101	10	CNS016LI	AL106896 Drosophil
c 35	50.4	5.2	1225	10	CNS0161D	AL106171 Drosophil
c 36	50	5.2	1092	10	CNS020K7	AL175696 Tetraodon
37	49.8	5.2	829	10	CG822206	CG822206 SOYDQ08TV
c 38	49.8	5.2	994	11	CNS04NOJ	AL298972 Tetraodon
c 39	49.6	5.2	770	9	CC919329	CC919329 t019n02ba
40	49.4	5.1	430	3	BI816150	BI816150 PfESToaa3
41	49.4	5.1	724	6	CA855370	CA855370 PfESToac3
42	49.4	5.1	838	10	AG869862	AG869862 Oryza sat
43	49.4	5.1	984	10	CNS006W8	AL066003 Drosophil
44	49.4	5.1	1029	10	CNS01ZGM	AL174271 Tetraodon
45	49.2	5.1	647	8	DT029073	DT029073 VVL030D03

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

```
  Query Match             8.2%;  Score 78.6;  DB 10;  Length 1101;
  Best Local Similarity   20.0%;  Pred. No. 2.1e-06;
  Matches  140;  Conservative  284;  Mismatches  271;  Indels    6;  Gaps    2;

Qy        131 ATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGATAAAATAAGAA 190
              ::::|: :: ::   :| : : |: :: | :|:  |:::| |  :|| :::::::::| ::
Db       1091 WDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWWWWWATWD 1032

Qy        191 ATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTCCAGAAATGTT 250
               ::::: ::|| |: : |:| :: |::| :  :::  |::::: |   :   :::  ::
Db       1031 TWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRRDGGRKR 972

Qy        251 TTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTAATAGTCAAAA 310
              ::: : ::  :: :||      |:  ::   : ::::  ::::::||:   |:::: | :
Db        971 KDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDK 912

Qy        311 ATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGATATGGTGTATG 370
              :::|: ::  :: :||:::::: : ::::: |:   :      ||:: :: : : :::|
Db        911 DDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTW 852

Qy        371 TAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACGTCATTCTACA 430
              :|||:: :  |: ::: : ::::: :|:  :||: : :|:   ::||| :  : :  :
Db        851 DAAADDWW--ADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDK 794

Qy        431 CAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATTTCTGGAATGA 490
              :||::::: | :||:| ||| : :  :  : :|:: :  : ::::    | | :
Db        793 RDAADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRK 734

Qy        491 TATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCGCAAACTACTG 550
              ::  : ::: :: |:::: :|::  ::::  |::| :| : :  |   : ::: |:
Db        733 WRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDRATWDRT 674

Qy        551 GAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGTGGAGAGTTCG 610
              :|::|:|  :| |: :|:::: ::::     :  :|||   :::   : :: |:::: :
Db        673 DAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTR 614

Qy        611 CAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGATATCATTTAT 670
              ::::::  :|: :|::|:::: :|:    ::  |:| :|:  :  :|: |    | |:|
Db        613 ADRWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTT----TATWT 558

Qy        671 ATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGAGCACC 730
              :||:  :: :: :| | || ||     : | |||  | |:||| |:   | : | |
Db        557 TWAARAAWWAWWAWATTTATWTTTWTTTWTTTTTTTTTTTTWTTTAAWWWAAWWTATWAAW 498

Qy        731 GTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATG 790
                :|  |||| | ||     |    |:    ::| :|: || : || ::|     |
Db        497 TAAAWAAAAAAAWAATTTTTTTTTTTTTWAWTAAAWWTAWTWTTTTWTTTTWWAATTTTTT 438

Qy        791 AAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGA 831
              :  |::||  |    :|||   | ||| |:::::|: | |
Db        437 WTWTTWWAATTATTTTWTTTWTAAAAAAAAWWWWWTWTATA 397


RESULT 2
CNS014PQ/c
LOCUS       CNS014PQ                 987 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN12P22 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL104456
VERSION     AL104456.1  GI:5616067
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .987
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN12P22"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             6.1%;  Score 59.2;  DB 10;  Length 987;
  Best Local Similarity   33.5%;  Pred. No. 0.031;
  Matches   90;  Conservative   63;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        685 AATAATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGAGCACCGTGTAAGTAAAAGA 744
              || ::: ::   : : | |||  | |||||| |   :          : ::  | ::: :
Db        957 AAAWWWWWWWWWWWWTTTTTTWTTTTTTTTTTTTTAAWWTTTAWTWWWWWHWTTTYHWWWW 898

Qy        745 GTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAA 804
               :|     || :   : ||: :: :: ::||  |: :::    ||  | |::||| |||
Db        897 TWTTTTWWATWWTTTWATTWTWWAWYCYWTATTTWTWWWAAWAATTTATAWWAAAAAAAA 838

Qy        805 AATTTTGCTACAAATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATA 864
              |::  :     ::|: |||| ||:: | :     ||:    |:|||   | |:||::| |
Db        837 AWWAAWMTNTAWWAWTATTAATTWWWATWAATWAATWAATTAWTTAATTAATWAAWWAAA 778

Qy        865 TATAAATATGGGCTTAGGTTTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAA 924
               ||  : ||    | |  || ||  |  :||::   :|:  | |   :|| |:|  ||||
Db        777 AATTTWAATTTAATAAWTTTCTTTAAAWWATWWTAAWAWTTTWTTTWWTTTTWTTAAAAA 718

Qy        925 ATAAAATATATTATAAAGACGGTTTTAAT 953
              | ||||:|::::| :|: |   :: ||:|
Db        717 AAAAAAWAWWWWAAWAWNANWTWWATAWT 689


RESULT 3
CNS00EVL
LOCUS       CNS00EVL                 1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
```

EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN
```

```
  Query Match             6.1%;  Score 58.6;  DB 10;  Length 1101;
  Best Local Similarity   34.4%;  Pred. No. 0.042;
  Matches  136;  Conservative   65;  Mismatches  194;  Indels    0;  Gaps    0;

Qy        542 AAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGTG 601
              :::  | |  | |||  |    ||  |    || |     || :   | :| |:| | |   :
Db        667 WWWTAATTAAATAATTTWAAATAWAAAAAAAAAAAAAAAAWATWAAWAATWATAWATAAWT 726

Qy        602 GAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGAT 661
              | | :    |  :|   ||:|::|||  ||: |     |   : |:::| :  :  ||
Db        727 TAAAAWAATAAAAWAAWAATWAWWATAATATWWATATATAWTTWTAWWWATWWAWWWTAT 786

Qy        662 ATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCCA 721
              |:   : |||:||  : : |    |:|    |   | |:   |  :  : | | ::   |
Db        787 AWATAWAATAWAAWAWAWATAAATAWATAWATWAAAWAWAWATAWWATWATATAWWAATA 846

Qy        722 GAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCA 781
              | |  |   :  |  |:||  |         |  |  :| ::|:|| |:: ::|  |  |
Db        847 WAWAAAAAAATWTAATATWAATWATAAWAAAAAATAWAWTTWWTWTTTWWAWWATATAAA 906

Qy        782 TAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGATA 841
              :|    :  |: | ||||  ||||   :   :  | || :|||||  ||     |    |:
Db        907 WAWATAWAAAWAAAAAAAAAAAAAATAAWAWWWTWATATTWTTATTAAAWTWTATWWATTW 966

Qy        842 GTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCTA 901
              |  : :|:|| |:||:  |:|||||| ::       ||  ||| |  |: |||||  ||
Db        967 ATTWWAWTWTATAWTTWTATWTATATATWWTAWTAWATATATTTATTAAWWTATATTTTA 1026

Qy        902 AAACTGTGCCTTTATTTAGAAAAATAAAATATATT 936
              :|| :    | |: | : : |:||| |: ::| |
Db       1027 WAAAWTAATATATWATAWWWTAWATATAWAWWAAT 1061
```

```
RESULT 4
CNS00EVL/c
LOCUS       CNS00EVL                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
```

please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"

ORIGIN

```
  Query Match             6.1%;  Score 58.6;  DB 10;  Length 1101;
  Best Local Similarity   32.6%;  Pred. No. 0.042;
  Matches  192;  Conservative   96;  Mismatches  300;  Indels    1;  Gaps    1;

Qy        365 TGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACGTCAT 424
              | | | |::  |    :  | | | : | ||::  |||| :  : :| | |     | :|
Db       1089 TWTTTTTWWTATWTTAWATATATAAWTAATTWWTWTATATWTAWWWTATWATATATTAWT 1030

Qy        425 TCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATTTCTG 484
              | :  | | || :: ||     ||| |   :  : | |  ||: |  |  | ||: : :
Db       1029 TTWTAAAATATAWWTTAATAAATATATWTAWTAWWATATATAWATAWAAWTATAWAWTWW 970

Qy        485 GAATGATATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCGCAAA 544
              | : ||:: ||: |:        |||  : :::    | | ||||    |     : |
Db        969 AATWAATWWATAW-AWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTTTTTWTTTWTA 911

Qy        545 CTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGTGGAG 604
              :  |  | | ::| : |   |:: |: :|     ||   ||:| |:| |     |
Db        910 TWTWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATATTAWATTTTT 851

Qy        605 AGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGATATC 664
              |:     :|||  :  :: ||:|: :|     : |   |  || :|    |  || :
Db        850 TWTWTATTWWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTWTWTTWTATTWT 791

Qy        665 ATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAG 724
              ||:||||:::  : |:::  :| :    |   :  ||  |:  : ||:||: |     :
Db        790 ATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTTWTTTTATTWT 731

Qy        725 AGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCATAG 784
              |    | |:  |:|     :       |      | |||:|  ||:      || |
Db        730 TTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTWTATTTWAAATTATTTAATT 671

Qy        785 GCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGATAGTG 844
              :   : :| |     | ||:|  |   |  | |||  |||| | : |      ||  |
Db        670 AWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAAAAATATTTTT 611

Qy        845 CATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCTAAAA 904
              |  ||  : :| |||||:|| |||||     |:    ||:|  || ||| :  | |||
Db        610 TTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTATTWWTTTAAA 551

Qy        905 CTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAAT 953
              | |   ||| ||||  |:|:: :|| : ||:|:::: |    |::::|
Db        550 TTATWTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWT 502
```

```
RESULT 5
CNS010D6
LOCUS       CNS010D6                1201 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN03E19 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098820
```

```
VERSION     AL098820.1  GI:5610431
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN03E19"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             6.0%;  Score 57.8;  DB 10;  Length 1201;
  Best Local Similarity   37.8%;  Pred. No. 0.063;
  Matches   74;  Conservative   50;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        758 CATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAA 817
              |||  |:|:|| :  ||: ::::: |  : |   : | ::|:::||:| ||:    | ||
Db        955 CATYATWAWTCABYATTHTYHHYMAAMAHTACCCWAANHMAMHYAARATTTWATAAAAAA 1014

Qy        818 ATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGC 877
              | || ||:: || |: ||  | | :    |: :|:| |||  ||:| ::|   |:|
Db       1015 AAAAATAKWRAAAARTTTTRGGTKDGTTTTWAWTWGTTTTTTTTDTTWRTTTTTRTTTTT 1074

Qy        878 TTAGGTTTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTA 937
              ::|: ::||| | : | | |   ::|  : |   :||::  : |||||||  | |||:|:
Db       1075 KKAKTKWTTTTTTKTTTTTTTTAWWATTKTTWWAWTTWWAAWAAAAAATAGTAAATAWTW 1134

Qy        938 TAAAGACGGTTTTAAT 953
              ::|:    |||:| ::
Db       1135 WWAWATGTGTTWTTWW 1150
```

```
RESULT 6
CNS00EPO/c
LOCUS       CNS00EPO               1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069493
VERSION     AL069493.1  GI:4949636
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29P01"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             6.0%;  Score 57.6;  DB 10;  Length 1101;
  Best Local Similarity   37.3%;  Pred. No. 0.069;
  Matches  123;  Conservative   61;  Mismatches  143;  Indels    3;  Gaps    1;

Qy        614 TTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGATATCATTTATATA 673
              :| |  :| :| : :|: | :: : |::     :: | |   ||| |:|:   : :|::|
Db        968 WTTTTKWTGWTTWGWTWTTAKDTRTTDWTWTTADKTAATTGAGTGAAWAKTTRWAWTWWA 909

Qy        674 AGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGAGCACCGTG 733
              : || |  |:|  :  |:     :| |::   |||||       |:|||       ||
Db        908 TAAKATTRTAGARTKTKTGTRRATRTDTTTDKAKAGTTTTGATRGAGAKAGATTTWTATG 849

Qy        734 TAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAG 793
              ||| ||| : ||::     || |  | :  : :| ||: |  : :  || :   |: ||:
Db        848 TAAATAATDTAGWWATAAAATTAAAAARAAWTWTTTTWATATWAWAAATTRTTAAWAAAR 789

Qy        794 ATAAAACCAAA---AATTTTGCTACAAATAATTATTTAAGAGATTGCGATAGTGCATTTA 850
              ::: ||  :||   |||| |   | ||:|| || ||  | | | |   |:  |  :|||
Db        788 WDRTAAAAWAATATAATTRTTAAAAAAWTATTTTTTATAAATAATTTWAWTWTTWWTTTT 729

Qy        851 TAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCTAAAACTGTGC 910
              |:|||  :: :| | |  |:||: :   ||  || |: |:  |||||  :|    | |
Db        728 TWGATAAWWAWAAAAAAKAWTAKTRARATAAATTATWATWATTATATATWATTTTTATTT 669

Qy        911 CTTTATTTAGAAAAATAAAATATATTATAA 940
              | |: :|| ::|  ||| | |||| :|||
Db        668 TTATWAWTAADWATTTAATAAATATAWTAA 639
```

```
RESULT 7
CNS00DKY
LOCUS       CNS00DKY               928 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
```

```
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .928
                      /organism="Drosophila melanogaster"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:7227"
                      /clone="BACR27A24"
                      /clone_lib="RPCI-98"
                      /note="end : T7"
ORIGIN

  Query Match             5.9%;  Score 57;  DB 10;  Length 928;
  Best Local Similarity   33.9%;  Pred. No. 0.092;
  Matches   96;  Conservative   62;  Mismatches  125;  Indels    0;  Gaps    0;

Qy        661 TATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCC 720
              ::|  ||:|:| |:  ||||:| |   |  : ||          ||: || || :|||
Db        538 WWTTTTTWAWAAAWTATTTWTTTTTTTTWTTTWAWTTTAATTWTATWTTAWTTAAAWW 597

Qy        721 AGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTC 780
                : |        |:| ||::| |   |      || |   : :||:|:  :  |||  |
Db        598 TAWWATTTTATAWTWAATAWWAAATATWTWATATAATATWTWTTWTWTAWTWTTATATAW 657

Qy        781 ATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGAT 840
              | |    |: || | :||   |||| |||    : ||| :|  | || |    ||    |
Db        658 AAAWAAAAWAAAAAAWAATAAAAAATTTTAWAAWTAAAAWAAAAATTTAWTTTTTTTTTT 717

Qy        841 AGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCT 900
              |    |: :| || :|| |||::::::||:| :|  :    :::: ::: :|:::
Db        718 TTTTWTATWTWAWATAWAAATATWWWWWWATWTWDGKNWNNNAWWWWWWWWWWAWWWWAA 777

Qy        901 AAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGA 943
              |::: : :   :  |   | |||||:|||| | ::::::::::
Db        778 AWWWWWWWWWWWAAAAAAAAAAAAAAWAAAAAAAWDDDDDDWDD 820


RESULT 8
CC264939/c
LOCUS       CC264939               1272 bp    DNA     linear   GSS 13-MAY-2003
DEFINITION  CH261-19L17_Sp6.1 CH261 Gallus gallus genomic clone CH261-19L17,
            genomic survey sequence.
ACCESSION   CC264939
VERSION     CC264939.1  GI:30612358
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 34
            High quality sequence stop: 261.
FEATURES              Location/Qualifiers
     source           1. .1272
                      /organism="Gallus gallus"
                      /mol_type="genomic DNA"
                      /strain="Red Jungle Fowl"
                      /db_xref="taxon:9031"
                      /clone="CH261-19L17"
                      /sex="female"
                      /cell_line="UCD001, inbred 256"
                      /clone_lib="CH261"
                      /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                      CH261 Female Chicken library - for library and clone
                      ordering information: http://www.chori.org/bacpac"
ORIGIN

  Query Match             5.7%;  Score 54.8;  DB 9;  Length 1272;
  Best Local Similarity   54.9%;  Pred. No. 0.28;
  Matches  107;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        759 ATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAA 818
              ||  | | |  ||  |||  | |||         | | | ||   | || |||  ||  ||
Db       1024 ATAATAAATNATAAATAAAATAATATATATAATAAAAATAAAATATAAATTTTTTATTAA 965

Qy        819 TAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGCT 878
              ||  |||||     |     |  |  || ||||  |||||| | ||||| ||||    |
Db        964 AAAAAATTTATATTTAAATTTTATAAATATATAATTTTAATAAAATATATATATAAATT 905

Qy        879 TAGGTTTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTAT 938
              ||      | |||  |||| | ||| | |   ||||||| |||| ||   |||  ||
Db        904 TAATAAAATATATTAATATAATTAAAATATTATTTTATTTAAAAAAAAAATTAATAAAAT 845

Qy        939 AAAGACGGTTTTAAT 953
              | | |   || ||||
Db        844 ATATATAATTATAAT 830


RESULT 9
AJ926878
LOCUS       AJ926878               1160 bp    mRNA    linear   EST 10-JUN-2005
DEFINITION  AJ926878 Theileria annulata merozoite Theileria annulata cDNA clone
            tam018g06_q1k, mRNA sequence.
ACCESSION   AJ926878
VERSION     AJ926878.1  GI:67497260
KEYWORDS    EST.
SOURCE      Theileria annulata
  ORGANISM  Theileria annulata
            Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
            Theileria.
REFERENCE   1  (bases 1 to 1160)
  AUTHORS   Pain,A., Renauld,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W.,
            Kerhornou,A., Aslett,M., Bishop,R., Bouchier,C., Cochet,M.,
            Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Fosker,N.,
            Gardner,M., Goble,A., Griffiths-Jones,S., Harris,D.E., Katzer,F.,
            Larke,N., Lord,A., Maser,P., McKellar,S., Mooney,P., Morton,F.,
            Nene,V., O'Neil,S., Price,C., Quail,M.A., Rabbinowitsch,E.,
            Rawlings,N.D., Rutter,S., Saunders,D., Seeger,K., Shah,T.,
            Squares,R., Squares,S., Tivey,A., Walker,A.R., Woodward,J.,
            Dobbelaere,D.A.E., Langsley,G., Rajandream,M.-A., McKeever,D.,
            Shiels,B., Tait,A., Barrell,B. and Hall,N.
  TITLE     The genome of the host-cell transforming parasite Theileria
            annulata and a comparison with T. parva
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Pain A
            The Pathogen Sequencing Unit
            The Wellcome Trust Sanger Institute
            Genome Campus, CB10 1SA, UNITED KINGDOM
            Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of
            Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
```

```
FEATURES             Location/Qualifiers
     source          1. .1160
                     /organism="Theileria annulata"
                     /mol_type="mRNA"
                     /isolate="Ankara (clone D7)"
                     /db_xref="taxon:5874"
                     /clone="tam018g06_q1k"
                     /dev_stage="merozoite"
                     /lab_host="Bos taurus (cow)"
                     /clone_lib="Theileria annulata merozoite"
                     /note="country: Turkey:Ankara"
ORIGIN

  Query Match             5.7%;  Score 54.6;  DB 1;  Length 1160;
  Best Local Similarity   41.0%;  Pred. No. 0.3;
  Matches  283;  Conservative    0;  Mismatches  398;  Indels    9;  Gaps    1;

Qy        256 TTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTAATAGTCAAAAATGGT 315
              |||| || ||     ||  |||   | |       | |   |  |||  |  ||      |
Db        354 TTATAAATTGTTTTTTAATTTTTATTNAAATTTANTAAAATTAAAATTNTTTAATTATTT 413

Qy        316 ATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGATATGGTGTATGTAAAT 375
              |||  ||  |  ||  ||  |     |       |   ||| |  |  |   | |||||
Db        414 TTTTAAAANTTNATTATTATTATNTTATATTAAATTAATTTATANNTTTTTATTATAAAT 473

Qy        376 TCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACGTCATTCTACACAAAT 435
              | |  |  |      | | |   |      ||  |||         | |||  | | | ||
Db        474 TAANNATAATNNNTTAATTNNNATATATATAATGAATAANTATANTTATTAAATATATAT 533

Qy        436 GATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATTTCTGGAATGATATGT 495
              | || || ||   |     |  | | |  |  || |||| |   |   |
Db        534 ATTATATATATTTANTANATNAATTATATAANTTATATATTAATTTATATNNTATATTTN 593

Qy        496 AAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCGCAAACTACTGGAAAA 555
              |  | |  |  |||   |||     |  |||||           || |
Db        594 ATTATATNTATTATAATTANTAAANAAAATTATATTTATTTAANTATNNTTATTTATNTT 653

Qy        556 TATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGTGGAGAGTTCGCAGTT 615
              |   || |  |||| |     | ||   |  | | |       |        ||      |
Db        654 TTATTTTTATATTTTATTATATATTTTTTAAAATTATATATAAATATTTTATTAATTANT 713

Qy        616 ATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGATATCATTTATATAAG 675
              |  |  |||| ||||| |  |   | ||| ||    |  | |||||||| |
Db        714 TTNATTTTTAATTTAATTTAATTTTTTNAATATATTTTTATTTATNNTTATTTATATTAT 773

Qy        676 GTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGAGCACCGTGTA 735
              |  ||| | ||  | |          ||  | || ||||      |  |     | ||
Db        774 ATAATTTTTTATTTTTTATTNTTTTAAATTTTTNTTATTTANAAATTTTANTAATATTTA 833

Qy        736 AGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGAT 795
              | ||| | ||       |       |||     ||   |||    |||    ||  |  |
Db        834 TATNAAATAATTATAAATTATAATANTTTTAATTTTAATAATNNAATAATATATTTATNT 893

Qy        796 AAAACCAAAA---------ATTTTGCTACAAATAATTATTTAAGAGATTGCGATAGTGCA 846
              || |  | ||         | |||  ||  ||||  || |     ||   ||| |  |
Db        894 AANATTATAATTATAATTTANTTTAATATTAATAAAATTTAATATTTTTATTATATTTAA 953

Qy        847 TTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCTAAAACT 906
              | |||  ||| |  | ||        ||  | | |    ||| ||  ||  |  | |   |
Db        954 TATATTTATTAANATTTAATATTTAATTAGATAAAAAATTTATAAATAATTATTTATTAT 1013

Qy        907 GTGCCTTTATTTAGAAAAATAAAATATATT 936
              |   ||| |||   ||  ||| | | ||||
Db       1014 ATTTTTTTTTTTTTAATTATATATTTTATT 1043


RESULT 10
CNS0058N
LOCUS       CNS0058N                994 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR11K13 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL057169
VERSION     AL057169.1  GI:4932557
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 994)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .994
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR11K13"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             5.6%;  Score 54.4;  DB 10;  Length 994;
  Best Local Similarity   42.9%;  Pred. No. 0.33;
  Matches  135;  Conservative   32;  Mismatches  145;  Indels    3;  Gaps    1;

Qy        639 TGAGCCAGATAATTCAGGTGGATATCATTTATATAAGGTTTTTTCTAATAATCTTGCGAC 698
              ||||||||| |   ||||  |    ||   |||  |       |  |  ||  |
Db        649 TGAGCCAGAAATCACAGGCTGTCGGCAAAAATAAGACAACAATCATTTTAGCCAATGCTA 708

Qy        699 AATCTTTGATGTTTTTTATCCCAGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGC 758
              | | | |  | | ||||   | | : |||| : |::  ::  |:|   : | ||  |:
Db        709 GGTATATAAAATATATTATAWTATATMMCACCCWATMMAAMWGGGRGGGGDAGGATTGGY 768

Qy        759 ATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAA 818
              |  :|: | ||  |:   || : |    |  || | |::|  ||||||  |:   | | :
Db        769 TTMAWTWATTTTTTTWTTATTTWAAAAAAAAAAAWAAAWWAAAAAAAATATKTAAAAATW 828

Qy        819 TAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGCT 878
              |||  |   || | |||   | | :  ||  |||   ||::  | | | |:  : |    |
Db        829 TAAAAAAAAAAAAAATTAWAAAAAWWTATAAATA---TTWWATTTTTNTWTTWTTTTTT 885

Qy        879 TAGGTTTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTAT 938
              |   ||||| ||: | |||  ||    | |   |||||:   |::| ::| :| | :: |
Db        886 TTTTTTTTTTTAWTTTTATTTTATTTTTTTTTTTTTATWATTAWWATWWATWTTTTWWKT 945

Qy        939 AAAGACGGTTTTAAT 953
              ||: |   |||||||
Db        946 AAWTATTTTTTTAAT 960


RESULT 11
CNS00DKY/c
LOCUS       CNS00DKY                928 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
```

```
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR27A24"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             5.6%;  Score 54.2;  DB 10;  Length 928;
  Best Local Similarity   32.7%;  Pred. No. 0.37;
  Matches   86;  Conservative   63;  Mismatches  114;  Indels    0;  Gaps    0;

Qy        681 TTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGAGCACCGTGTAAGTAA 740
              ||::::::::::: :     | |:|  | ||||||  |   : : :  :   : ::  :::
Db        826 TTMWWHHHWHHHHHHWTTTTTTTWTTTTTTTTTTTTTWWWWWWWWWWWTTTWWWWTWWWW 767

Qy        741 AAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAAC 800
              :: : :|     : ::  || :::::   |  :||  |: :      |  || | ||||
Db        766 WWWWWWTNNNWNMCHWAWATWWWWWWATATTTWTATWTWAWATAWAAAAAAAAAAAAAAW 707

Qy        801 CAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAAT 860
              |||  |||:  |  :  : |  ||||   |  :|    |: |  :|||:|| ||  :|:
Db        706 TAAATTTTTWTTTTAWTTWTAAAATTTTTTATTWTTTTTTWTTTTWTTTWTATATAAWAW 647

Qy        861 TATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAG 920
              ||:|:| :|:        |:   | || ::| | :|:  |||||  : |   |||| :||
Db        646 TAWAWAAWAWATATTATATWAWATATTTWWTATTWAWTATAAAATWWTAWWTTTAAWTAA 587

Qy        921 AAAAATAAAATATATTATAAAGA 943
              :| |: |  | | :|:| |:| |
Db        586 WATAWAATTAAAAWTWAAAWAAA 564


RESULT 12
CNS004ZW
LOCUS       CNS004ZW               1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR11E08 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL055440
VERSION     AL055440.1  GI:4932241
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR11E08"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             5.6%;  Score 54;  DB 10;  Length 1101;
  Best Local Similarity   38.8%;  Pred. No. 0.41;
  Matches  127;  Conservative   46;  Mismatches  153;  Indels    1;  Gaps    1;

Qy        612 AGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGATATCATTTATA 671
              |: | |  | | ||| | |: |      |      | ::  |  |  |  |  | |::||:|
Db        714 AKATTTNTAAAKTAAAAAAAWATWTATATWTMWAATWWTTTAWATWTAAAATTWAWWTAWA 773

Qy        672 TAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGAGCACCG 731
              ||    |:: | |: :  : ::          ||| || |||| |||   |:| |: |
Db        774 TATATATWWNTATWTWTTWWWWTTATNTTATTTTTATATTTTATATATTAASAAAVAATAT 833

Qy        732 TGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGA 791
              : ||| |:||| | |       ||      |      || ||  |:||  | ||     ||
Db        834 WTTAAATMAAATANTATANTTATATTATTTAAATTTATTCATWAATATAATTTTTAATAT 893

Qy        792 AGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGATAGTGCATTTAT 851
              :  :||  ||: |    |  |  : ||| :| |:||:    :|   :|  ::  ::|||
Db        894 TAWATMAAATAAWTAAAATWTTTAWTATATWTTTWTAWTTTTKTTATWTTAWDATWWTAT 953

Qy        852 AGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCTAAAACTGTGCC 911
              || |:  |||::|||||  |:    :|:  : ||| ||| |:||:   |||| | |
Db        954 TWATATWTATATWWATAAA-AWWTAAWTWWAWATTTATATATWTAWTAAAAAAWTATTTW 1012

Qy        912 TTTATTTAGAAAAATAAAATATATTAT 938
              | | ||   |:|||  |: :| ||::|
Db       1013 TRTTTTGTTAWAAAATAWTWAAATWWT 1039


RESULT 13
BH250178
LOCUS       BH250178               751 bp    DNA     linear   GSS 26-NOV-2001
DEFINITION  BOGAF63TF BOGA Brassica oleracea genomic clone BOGAF63, genomic
            survey sequence.
ACCESSION   BH250178
VERSION     BH250178.1  GI:17073660
KEYWORDS    GSS.
```

```
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
            Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea and its
            application to gene discovery and annotation in Arabidopsis
  JOURNAL   Genome Res. 15 (4), 487-495 (2005)
   PUBMED   15805490
COMMENT     Other_GSSs: BOGAF63TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .751
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGAF63"
                     /clone_lib="BOGA"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             5.6%;  Score 53.8;  DB 9;  Length 751;
  Best Local Similarity   48.8%;  Pred. No. 0.44;
  Matches  145;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy        206 TAAATGAAGAAAACTTAGGAATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATG 265
              ||| || | ||| |||  |||   |   |  ||   ||    |||   | | || |
Db        188 TAATTGCATAAATCTTTTGAAACAAATAAGATATCTTTAACATATTTAGATCAACATAAT 247

Qy        266 GAAAATATCTTTGGATGATTGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGG 325
              |   ||| || | |||||||    |||| || ||||||  |||||| | ||     |
Db        248 AATTTTATTTTAGTATGATTGATTGTGATTATATAATAGATAAAAATAGGATAGAATATT 307

Qy        326 TTTTTTCGATATTAAAGTCCCGGCCTGCATTAGATATGGTGTATGTAAATTCAGCAGCAA 385
              | ||||  || ||| |  |      ||| ||     ||| |||||   | |||    | | |
Db        308 TATTTTTCATTTTATAAACGATAACTGAATATATTATTGTGTAATAATATTTCAAAACTA 367

Qy        386 AGACTGAGTTAAACTATAATGCTGATGTGAGGACGTCATTCTACACAAATGATGTAGATT 445
              |     |   | ||| | |  |       |  |  | | | ||  | ||
Db        368 ATTACAAAACTAGTGAAAATATTTACATATAATTTTTTTAATTAAGATCTTATTAAAATC 427

Qy        446 TTATTTCAGACGTGAAAGTTATGTTCACGTTTATTTCTGGAATGATATGTAAGAAAA 502
              ||  |  |||  |       || ||||   |  ||      | ||||| || || |
Db        428 TTTATAAAGAGATTTTCAAAATTTTCAATATATATTTATATTTTATATGAAAAAATA 484


RESULT 14
CNS04D0K/c
LOCUS       CNS04D0K                945 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            101H21 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL285149
VERSION     AL285149.1  GI:8023560
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
   PUBMED   10899143
REFERENCE   3  (bases 1 to 945)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .945
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="101H21"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG101CD11LP1
                     end : T7"
ORIGIN

  Query Match             5.5%;  Score 53.4;  DB 11;  Length 945;
  Best Local Similarity   38.0%;  Pred. No. 0.54;
  Matches  167;  Conservative   46;  Mismatches  224;  Indels    3;  Gaps    1;

Qy        510 AATTGTCAAAGCCGTTGGTATTTTCAGTCCGCAAACTACTGGAAAATATCTTATGCATTT 569
              | || | | :    ||   | ||| | :     |: || |    |||:| ||    |||
Db        851 ATTTATTANWTANWTTATAANTTTTATWNTAATAWATAATWTTTAATWTTTTTATTATTA 792

Qy        570 AACATGGCAATTGCCATTACTTAAACAGGGTGGAGAGTTCGCAGTTATCCATAATAATAT 629
              :| ||     |    ||:: |: |           | ||     ||:|     ||| | |
Db        791 WAAATTTTTTATTTATWWATWTATTTTATATTTAATTTATATTTTWTATTATATATTTT 732

Qy        630 AATTGAGGCTGAGCCAGATAATTCAGGTGGATATCATTTATATAAGGTTTTTTCTAATAA 689
              ||     |    :  |::|| :  |  || | |||| ||| :  |||| | |  ||
Db        731 TTTTTNTWTTATWAWWTTWWTTWTTTTTATTTTATTTTTATTWTWTTTTATTTTTTAT 672

Qy        690 TCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGAGCACCGTGTAAGTAAAAGAGTTCG 749
              |:     ::: |:: |: :::|:|||     : |  |   | |   |       ||
Db        671 ATTW---TAWWWTTWWTAWWWWWTWTATWTWTTWTATAAATATATTTATTTTTATWTTTA 615

Qy        750 CGCATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTT 809
                | |  || |||:|| |||      || |||     :  | ||||||   | :|
Db        614 TTWTTAAAAATATTTWTTTTTAAAATTTTTTATATTATTWTTAAATAAAAAATATWAAAA 555

Qy        810 TGCTACAAATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAA 869
              |  |: |:|:|||: | | | :  ||   | |    :|:|    |||| ||:|:| | :
Db        554 TWWTWTAWAWAATWTTATTATWTTTTTTTAAAAAWWWTWTTATTTTTTATTTWTWTTTTW 495

Qy        870 ATATGGGCTTAGGTTTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAA 929
              |:|     |||  ||||  ||  ||||     |||:   :    :  |    |||||  ||
Db        494 AWAATWNTTTAAWTTTTATTAATTATAAAAAAAAWAAWWWTAAWRNTANTAAAAAAAAAN 435

Qy        930 ATATATTATAAAGACGGTTT 949
              : |: :| |    |  | ||
Db        434 WAADKKTGTNKTAAATGNTT 415
```

```
RESULT 15
BH955189
LOCUS       BH955189               706 bp    DNA     linear   GSS 01-OCT-2002
DEFINITION  odi49d09.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BH955189
VERSION     BH955189.1  GI:23436416
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 706)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
            Nash,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: odi49  row: d  column: 09
            Seq primer: -28RPpOT reverse
            Class: shotgun
            High quality sequence start: 16
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1. .706
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
ORIGIN

  Query Match             5.5%;  Score 53.2;  DB 9;  Length 706;
  Best Local Similarity   47.1%;  Pred. No. 0.59;
  Matches  197;  Conservative    0;  Mismatches  218;  Indels    3;  Gaps    1;

Qy        529 ATTTTCAGTCCGCAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTA 588
              | ||   ||    | |  | | |||| ||  || |   |||   |||  ||||    |
Db        218 AGTTGTTGTTTATATATGATTTGAAAGTAAATTTTTGGTTTGTTATGTTAATTTGTTGTT 277

Qy        589 CTTAAACAGGGTGGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGAT 648
               || ||     |  ||  |     | |||| | |||  | |  |||  |     |   ||
Db        278 TTTGAAATATATAAAGGTTAAATTGCTATCTAAAATGTTTTTTTTGTTGAATTTCTGTAT 337

Qy        649 AATTCAGGTGGATATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGAT 708
              ||   | ||   ||  || ||||||   | | | ||   | |    |   ||| |
Db        338 AACAAACGTTTTTAATATATATATATATATATATATATATTTATGTTATCTTATTTTAAA 397

Qy        709 GTTTTTTATCCCAGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTC 768
              |||  |   | | |  |     ||    | |    |       |    |   || ||
Db        398 AACGTTTTTAATAAATAAAAATAAATATAGCATATTTATAATATTTTGAAAACGTTTTTA 457

Qy        769 TTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTAT--- 825
               ||| ||  || || |       | |||  |  | | || |   || |||  ||| |
Db        458 ATACATATATTTATTGTGTTATCATATATTAATATATATATATATATAAACGATTTTATA 517

Qy        826 TTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTT 885
              || |     |   | |     |||| | || ||  |||||||| ||||    ||| ||||
Db        518 TTTATTATATATTAAAAACGTTTTAAATATATATATATATATATATATATTTATGTTT 577

Qy        886 TTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGA 943
                | ||| |||||   |||  | |    | | | | | ||||| |  |||||   || |
Db        578 ATATATATATATTTAAAACGTTTTTAATAAATAAAATAAATATAGCATATTTATAATA 635


Search completed: May 12, 2006, 03:29:23
Job time : 3017.39 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 19:15:19 ; Search time 131.709 Seconds
(without alignments)
12996.800 Million cell updates/sec

Title: US-09-423-093B-3_COPY_2352_3314
Perfect score: 963
Sequence: 1 atgagtaaaaaacctcttct.....cggttttaatgcggaaataa 963

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	54.4	5.6	1141	3	US-09-806-708B-22	Sequence 22, Appl
	2	50.6	5.3	906	3	US-09-495-406-14	Sequence 14, Appl
	3	50.6	5.3	906	3	US-09-816-028A-26	Sequence 26, Appl
	4	50.6	5.3	906	3	US-10-303-162-26	Sequence 26, Appl
	5	50.6	5.3	906	3	US-10-303-134-26	Sequence 26, Appl
	6	50.6	5.3	906	3	US-10-303-118-26	Sequence 26, Appl
	7	50.6	5.3	906	3	US-10-303-128-26	Sequence 26, Appl
c	8	50.6	5.3	11474	3	US-09-495-406-1	Sequence 1, Appli
c	9	50.6	5.3	11474	3	US-09-816-028A-1	Sequence 1, Appli
c	10	50.6	5.3	11474	3	US-10-303-162-1	Sequence 1, Appli
c	11	50.6	5.3	11474	3	US-10-303-134-1	Sequence 1, Appli
c	12	50.6	5.3	11474	3	US-10-303-118-1	Sequence 1, Appli
c	13	50.6	5.3	11474	3	US-10-303-128-1	Sequence 1, Appli
c	14	48.6	5.0	1141	3	US-09-806-708B-22	Sequence 22, Appl
	15	46.4	4.8	36242	3	US-09-949-016-12996	Sequence 12996, A
	16	46.4	4.8	36242	3	US-09-949-016-12997	Sequence 12997, A
	17	46.4	4.8	36242	3	US-09-949-016-12998	Sequence 12998, A
	18	46.4	4.8	36242	3	US-09-949-016-12999	Sequence 12999, A
	19	46.4	4.8	36242	3	US-09-949-016-13000	Sequence 13000, A
	20	46	4.8	1356	3	US-09-543-681A-3229	Sequence 3229, Ap
	21	45.6	4.7	30820	3	US-09-949-016-17145	Sequence 17145, A
	22	45.4	4.7	96922	3	US-09-949-016-17061	Sequence 17061, A
	23	45	4.7	211049	3	US-09-949-016-15770	Sequence 15770, A
c	24	44.4	4.6	19732	3	US-09-949-016-12870	Sequence 12870, A
c	25	44.4	4.6	19732	3	US-09-949-016-14923	Sequence 14923, A
c	26	44.4	4.6	19733	3	US-09-949-016-14336	Sequence 14336, A
	27	43.8	4.5	50000	3	US-09-662-254B-24	Sequence 24, Appl
c	28	43.6	4.5	16573	3	US-09-949-016-14876	Sequence 14876, A
	29	43.6	4.5	114139	3	US-09-949-016-16536	Sequence 16536, A
	30	42.8	4.4	601	3	US-09-949-016-30530	Sequence 30530, A
	31	42.8	4.4	601	3	US-09-949-016-30531	Sequence 30531, A
	32	42.8	4.4	601	3	US-09-949-016-37149	Sequence 37149, A
	33	42.8	4.4	601	3	US-09-949-016-37150	Sequence 37150, A
	34	42.8	4.4	601	3	US-09-949-016-37163	Sequence 37163, A
	35	42.8	4.4	601	3	US-09-949-016-37164	Sequence 37164, A
	36	42.8	4.4	601	3	US-09-949-016-145867	Sequence 145867,
	37	42.8	4.4	601	3	US-09-949-016-145868	Sequence 145868,
	38	42.8	4.4	601	3	US-09-949-016-146135	Sequence 146135,
	39	42.8	4.4	601	3	US-09-949-016-146136	Sequence 146136,
	40	42.8	4.4	601	3	US-09-949-016-146403	Sequence 146403,
	41	42.8	4.4	601	3	US-09-949-016-146404	Sequence 146404,
c	42	42.8	4.4	205044	3	US-09-949-016-15851	Sequence 15851, A
c	43	42.8	4.4	205044	3	US-09-949-016-15852	Sequence 15852, A
c	44	42.8	4.4	205044	3	US-09-949-016-15853	Sequence 15853, A
c	45	42.8	4.4	223471	3	US-09-949-016-12387	Sequence 12387, A

ALIGNMENTS

RESULT 1
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 5.6%; Score 54.4; DB 3; Length 1141;
Best Local Similarity 12.2%; Pred. No. 0.0005;
Matches 81; Conservative 260; Mismatches 314; Indels 7; Gaps 2;

```
Qy      270 ATATCTTTGGATGATTGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTT 329
            :| :  :|| ::: :  :   : :: :::||:   :  : ::::::  |   : :    |
Db      114 MTNKWGKTGWRHRYWRWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNAT 173

Qy      330 TTCGATATTAAAGTCCCGGCCTGCATTAGATATGGTGTATGTAAATTCAGCAGCAAAGAC 389
              : ::| ::::    |: :   :  :|:::: ::: :  :::  : :::    ::::::
Db      174 GWDDDTKYHMWNNNGCBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDM 233

Qy      390 TGAGTTAAACTATAATGCTGATGTGAGGACGTCATTCTACACAAATGATGTAGATTTTAT 449
             :::  ::: |: ::  : :|::| :::::   :|:  |:: : :       |:|  : |:
Db      234 CKRKVRRWVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRAK 293

Qy      450 TTCAGACGTGAAAGTTATGTTCACGTTTATTTCTGGAATGATATGTAAGAAAACTGATGC 509
            :: : : :      |:  :|| :    :|     |:| ::::::: | : :
Db      294 YWGWNRABVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNN 353

Qy      510 AATTGTCAAAGCCGTTGGTATTTTCAGTCCGCAAACTACTGGAAAATATCTTATGCATTT 569
            :|::: ::    ::    |:::::|   : ::   | :::: ::      ||::: :: :
Db      354 NWTYKKARHBARWDWVWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAW 413
```

```
Qy        570 AACATGGCAATTGCCATTACTTAAACAGGGTGGAGAGTTCGCAGTTATCCATAATAATAT 629
Db        414 YWKMDMDWBGTYNNNNNGGRTYY-----GWTKNKKMWTYYKWKANNCKWRAWDHKTCTHN 468

Qy        630 AATTGAGGCTGAGCCAGATAATTCAGGTGGATATCATTTATATAAGGTTTTTTCTAATAA 689
Db        469 NTTWWKMKTYWNNCYWKSMTNGKSHRBAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWY 528

Qy        690 TCTTGCGACAATCTTTGATGTTTTTTATCCCAGAGAGCACCGTGTAAGTAAAAGAGTTCG 749
Db        529 KYBVCSKWWNNYAAWY--TKSSWNYTSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYG 586

Qy        750 CGCATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTT 809
Db        587 YKWNTRWWBWSHTWBHBRAGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNN 646

Qy        810 TGCTACAAATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAA 869
Db        647 NNNNNNNNATCARDDYYAASRWYAMANAKWYYYKBAANNAYYTHANNWWGCWNNATDTRR 706

Qy        870 ATATGGGCTTAGGTTTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAA 929
Db        707 TMWKNNNNNNAGTWKNNNNNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADAAA 766

Qy        930 AT 931
Db        767 BT 768
```

RESULT 2
US-09-495-406-14
; Sequence 14, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000110US
; CURRENT APPLICATION NUMBER: US/09/495,406
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(906)
; OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
; OTHER INFORMATION: OH4384 (ORF 6a of LOS biosynthesis locus)
US-09-495-406-14

Query Match 5.3%; Score 50.6; DB 3; Length 906;
Best Local Similarity 49.1%; Pred. No. 0.0042;
Matches 134; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359
```

RESULT 3
US-09-816-028A-26
; Sequence 26, Application US/09816028A
; Patent No. 6699705
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(906)
; OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
; OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
; OTHER INFORMATION: biosynthesis locus)
US-09-816-028A-26

Query Match 5.3%; Score 50.6; DB 3; Length 906;
Best Local Similarity 49.1%; Pred. No. 0.0042;
Matches 134; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359
```

RESULT 4
US-10-303-162-26
; Sequence 26, Application US/10303162
; Patent No. 6723545
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.

```
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,162
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-303-162-26

  Query Match             5.3%;  Score 50.6;  DB 3;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 0.0042;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |      |      |  || || |  |  || ||||       ||| |   | | |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |    |||||||    | |    |||||  |  |     ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 5
US-10-303-134-26
; Sequence 26, Application US/10303134
; Patent No. 6825019
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,134
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-303-134-26

  Query Match             5.3%;  Score 50.6;  DB 3;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 0.0042;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |      |      |  || || |  |  || ||||       ||| |   | | |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |    |||||||    | |    |||||  |  |     ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 6
US-10-303-118-26
; Sequence 26, Application US/10303118
; Patent No. 6905867
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,118
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-303-118-26

  Query Match             5.3%;  Score 50.6;  DB 3;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 0.0042;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;
```

```
Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |     ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
                |       |   || || |  |  || ||||       ||| |   | |  |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |    |||||||    | |    |||| | |    ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 7
US-10-303-128-26
; Sequence 26, Application US/10303128
; Patent No. 6911337
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,128
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-303-128-26

  Query Match             5.3%;  Score 50.6;  DB 3;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 0.0042;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |     ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
                |       |   || || |  |  || ||||       ||| |   | |  |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |    |||||||    | |    |||| | |    ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 8
US-09-495-406-1/c
; Sequence 1, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000110US
;  CURRENT APPLICATION NUMBER: US/09/495,406
;  CURRENT FILING DATE:  2000-01-31
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 11474
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
;   OTHER INFORMATION: including LOS biosynthesis locus
US-09-495-406-1

  Query Match             5.3%;  Score 50.6;  DB 3;  Length 11474;
  Best Local Similarity   49.1%;  Pred. No. 0.0095;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |     ||  |||||
Db       5873 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 5814

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db       5813 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 5754

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
                |       |   || || |  |  || ||||       ||| |   | |  |
Db       5753 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 5694

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |    |||||||    | |    |||| | |    ||  | |    || ||| |
Db       5693 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 5634

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db       5633 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 5601


RESULT 9
US-09-816-028A-1/c
; Sequence 1, Application US/09816028A
; Patent No. 6699705
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/09/816,028A
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
```

```
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 11474
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
;   OTHER INFORMATION: including LOS biosynthesis locus
US-09-816-028A-1

  Query Match             5.3%;  Score 50.6;  DB 3;  Length 11474;
  Best Local Similarity   49.1%;  Pred. No. 0.0095;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db       5873 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 5814

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db       5813 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAACTTAGG 5754

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |     |       |  || || |  |  || ||||       ||| |   | |  |
Db       5753 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 5694

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |    |||||||    | |    ||||| | |    ||  | |    || ||| |
Db       5693 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 5634

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db       5633 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 5601


RESULT 10
US-10-303-162-1/c
; Sequence 1, Application US/10303162
; Patent No. 6723545
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,162
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 11474
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
;   OTHER INFORMATION: including LOS biosynthesis locus
US-10-303-162-1

  Query Match             5.3%;  Score 50.6;  DB 3;  Length 11474;
  Best Local Similarity   49.1%;  Pred. No. 0.0095;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db       5873 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 5814

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db       5813 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAACTTAGG 5754

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |     |       |  || || |  |  || ||||       ||| |   | |  |
Db       5753 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 5694

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |    |||||||    | |    ||||| | |    ||  | |    || ||| |
Db       5693 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 5634

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db       5633 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 5601


RESULT 11
US-10-303-134-1/c
; Sequence 1, Application US/10303134
; Patent No. 6825019
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,134
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 11474
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
;   OTHER INFORMATION: including LOS biosynthesis locus
US-10-303-134-1

  Query Match             5.3%;  Score 50.6;  DB 3;  Length 11474;
  Best Local Similarity   49.1%;  Pred. No. 0.0095;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db       5873 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 5814

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db       5813 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAACTTAGG 5754

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |     |       |  || || |  |  || ||||       ||| |   | |  |
Db       5753 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 5694

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |    |||||||    | |    ||||| | |    ||  | |    || ||| |
Db       5693 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 5634

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db       5633 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 5601
```

```
RESULT 12
US-10-303-118-1/c
; Sequence 1, Application US/10303118
; Patent No. 6905867
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,118
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 11474
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
;   OTHER INFORMATION: including LOS biosynthesis locus
US-10-303-118-1

  Query Match             5.3%;  Score 50.6;  DB 3;  Length 11474;
  Best Local Similarity   49.1%;  Pred. No. 0.0095;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |     ||  |||||
Db       5873 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 5814

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db       5813 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 5754

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |      |      |  || || |  |  || ||||        ||| |   | |  |
Db       5753 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 5694

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |   |||||||    | |    |||| | |   |  || | |    || |||  |
Db       5693 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 5634

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db       5633 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 5601


RESULT 13
US-10-303-128-1/c
; Sequence 1, Application US/10303128
; Patent No. 6911337
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,128
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 11474
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
;   OTHER INFORMATION: including LOS biosynthesis locus
US-10-303-128-1

  Query Match             5.3%;  Score 50.6;  DB 3;  Length 11474;
  Best Local Similarity   49.1%;  Pred. No. 0.0095;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |     ||  |||||
Db       5873 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 5814

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db       5813 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 5754

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |      |      |  || || |  |  || ||||        ||| |   | |  |
Db       5753 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 5694

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |   |||||||    | |    |||| | |   |  || | |    || |||  |
Db       5693 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 5634

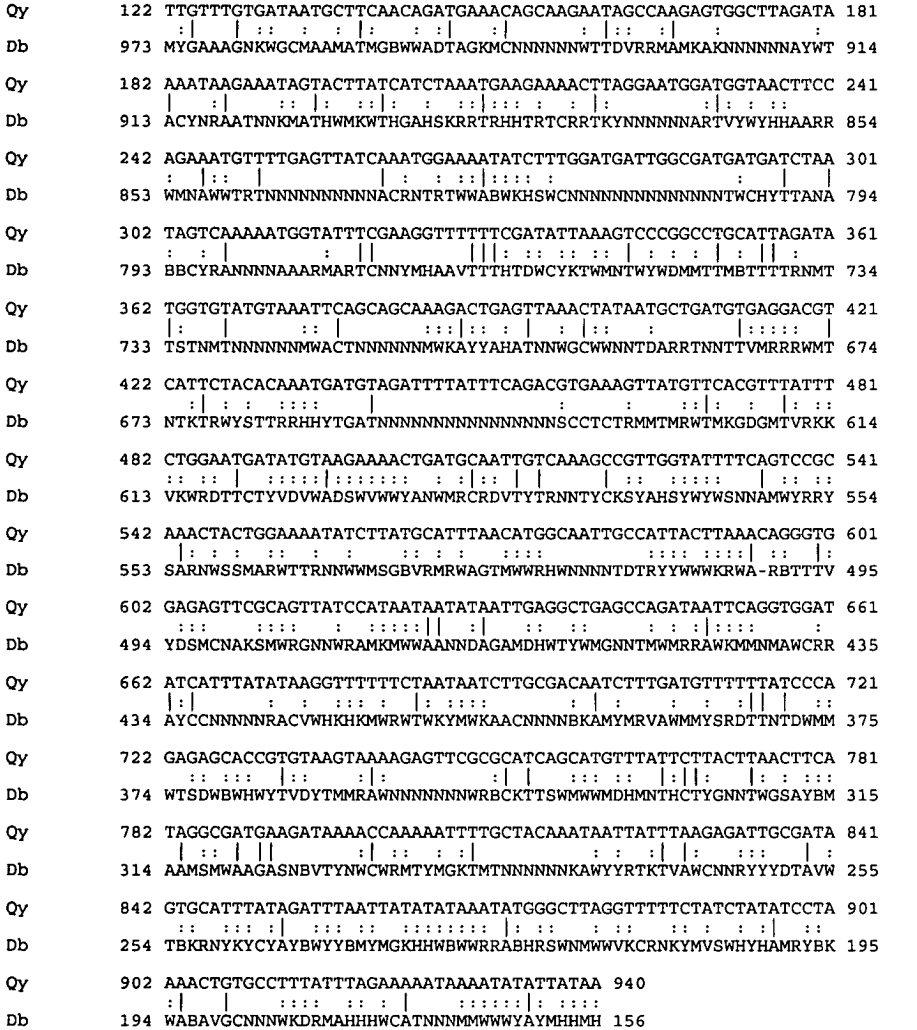
Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db       5633 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 5601


RESULT 14
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             5.0%;  Score 48.6;  DB 3;  Length 1141;
  Best Local Similarity   11.2%;  Pred. No. 0.014;
  Matches  105;  Conservative  321;  Mismatches  512;  Indels    1;  Gaps    1;

Qy          2 TGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTGG 61
              | : | :| :    |:: ::  |    | :|: ::   |:  |  :|  ::::: ::
Db       1093 TAWWTNHAKRGATMCWYWYWTGTNRRWCMRTYAMRTWYTRSNANWSCATKBMWWTMKWYA 1034

Qy         62 CTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTTA 121
               :::| :   : |::    : |     : |      ::  | :      |      ||   :
Db       1033 TKYRTAWYAMWCAWRNNNMWCATNGYAKSCATNNAMWYATTRWAAYAAAKWARWAGNNMR 974
```



RESULT 15
US-09-949-016-12996
; Sequence 12996, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12996
; LENGTH: 36242
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12996

Query Match 4.8%; Score 46.4; DB 3; Length 36242;
Best Local Similarity 50.4%; Pred. No. 0.16;
Matches 113; Conservative 0; Mismatches 111; Indels 0; Gaps 0;



Search completed: May 12, 2006, 03:51:04
Job time : 136.709 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:31:18 ; Search time 616.595 Seconds
(without alignments)
12915.142 Million cell updates/sec

Title: US-09-423-093B-3_COPY_2352_3314
Perfect score: 963
Sequence: 1 atgagtaaaaaacctcttct.....cggttttaatgcggaaataa 963

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	963	100.0	12441	3	US-09-423-093A-3	Sequence 3, Appli
	2	50.6	5.3	906	3	US-09-816-028A-26	Sequence 26, Appl
	3	50.6	5.3	906	6	US-10-303-161-26	Sequence 26, Appl
	4	50.6	5.3	906	6	US-10-303-118-26	Sequence 26, Appl
	5	50.6	5.3	906	6	US-10-303-128-26	Sequence 26, Appl
	6	50.6	5.3	906	6	US-10-303-134-26	Sequence 26, Appl
	7	50.6	5.3	906	6	US-10-303-162-26	Sequence 26, Appl
	8	50.6	5.3	906	8	US-10-735-419-26	Sequence 26, Appl
	9	50.6	5.3	906	8	US-10-820-536-26	Sequence 26, Appl
	10	50.6	5.3	906	8	US-10-845-408-26	Sequence 26, Appl
	11	50.6	5.3	906	8	US-10-845-412-26	Sequence 26, Appl
	12	50.6	5.3	906	8	US-10-846-219-26	Sequence 26, Appl
	13	50.6	5.3	906	8	US-10-821-604-26	Sequence 26, Appl
	14	50.6	5.3	906	8	US-10-847-983-26	Sequence 26, Appl
	15	50.6	5.3	906	8	US-10-821-573-26	Sequence 26, Appl
	16	50.6	5.3	906	8	US-10-850-807-26	Sequence 26, Appl
	17	50.6	5.3	906	8	US-10-850-125-26	Sequence 26, Appl
	18	50.6	5.3	906	8	US-10-830-825-26	Sequence 26, Appl
	19	50.6	5.3	906	9	US-10-962-334-26	Sequence 26, Appl
	20	50.6	5.3	906	9	US-10-830-997-26	Sequence 26, Appl
	21	50.6	5.3	906	9	US-10-962-235-26	Sequence 26, Appl
	22	50.6	5.3	906	9	US-10-961-882-26	Sequence 26, Appl
c	23	50.6	5.3	11474	3	US-09-816-028A-1	Sequence 1, Appli
c	24	50.6	5.3	11474	6	US-10-303-161-1	Sequence 1, Appli
c	25	50.6	5.3	11474	6	US-10-303-118-1	Sequence 1, Appli
c	26	50.6	5.3	11474	6	US-10-303-128-1	Sequence 1, Appli
c	27	50.6	5.3	11474	6	US-10-303-134-1	Sequence 1, Appli
c	28	50.6	5.3	11474	6	US-10-303-162-1	Sequence 1, Appli
c	29	50.6	5.3	11474	8	US-10-735-419-1	Sequence 1, Appli
c	30	50.6	5.3	11474	8	US-10-820-536-1	Sequence 1, Appli
c	31	50.6	5.3	11474	8	US-10-845-408-1	Sequence 1, Appli
c	32	50.6	5.3	11474	8	US-10-845-412-1	Sequence 1, Appli
c	33	50.6	5.3	11474	8	US-10-846-219-1	Sequence 1, Appli
c	34	50.6	5.3	11474	8	US-10-821-604-1	Sequence 1, Appli
c	35	50.6	5.3	11474	8	US-10-847-983-1	Sequence 1, Appli
c	36	50.6	5.3	11474	8	US-10-821-573-1	Sequence 1, Appli
c	37	50.6	5.3	11474	8	US-10-850-807-1	Sequence 1, Appli
c	38	50.6	5.3	11474	8	US-10-850-125-1	Sequence 1, Appli
c	39	50.6	5.3	11474	8	US-10-830-825-1	Sequence 1, Appli
c	40	50.6	5.3	11474	9	US-10-962-334-1	Sequence 1, Appli
c	41	50.6	5.3	11474	9	US-10-830-997-1	Sequence 1, Appli
c	42	50.6	5.3	11474	9	US-10-962-235-1	Sequence 1, Appli
c	43	50.6	5.3	11474	9	US-10-961-882-1	Sequence 1, Appli
c	44	48.6	5.0	633	4	US-09-925-065A-464691	Sequence 464691,
	45	48.4	5.0	5654	6	US-10-311-455-1847	Sequence 1847, Ap

ALIGNMENTS

RESULT 1
US-09-423-093A-3
; Sequence 3, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12441
; TYPE: DNA
; ORGANISM: Salmonella enterica
US-09-423-093A-3

Query Match 100.0%; Score 963; DB 3; Length 12441;
Best Local Similarity 100.0%; Pred. No. 5.2e-195;
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2352 ATGAGTAAAAAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTG 2411

Qy         61 GCTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2412 GCTCGTTTACTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTT 2471

Qy        121 ATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2472 ATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGAT 2531

Qy        181 AAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2532 AAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTC 2591

Qy        241 CAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2592 CAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTA 2651
```

```
Qy        301 ATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2652 ATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGAT 2711

Qy        361 ATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2712 ATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATAATGCTGATGTGAGGACG 2771

Qy        421 TCATTCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2772 TCATTCTACACAAATGATGTAGATTTTATTTCAGACGTGAAAGTTATGTTCACGTTTATT 2831

Qy        481 TCTGGAATGATATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2832 TCTGGAATGATATGTAAGAAAACTGATGCAATTGTCAAAGCCGTTGGTATTTTCAGTCCG 2891

Qy        541 CAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2892 CAAACTACTGGAAAATATCTTATGCATTTAACATGGCAATTGCCATTACTTAAACAGGGT 2951

Qy        601 GGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2952 GGAGAGTTCGCAGTTATCCATAATAATATAATTGAGGCTGAGCCAGATAATTCAGGTGGA 3011

Qy        661 TATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3012 TATCATTTATATAAGGTTTTTTCTAATAATCTTGCGACAATCTTTGATGTTTTTTATCCC 3071

Qy        721 AGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3072 AGAGAGCACCGTGTAAGTAAAAGAGTTCGCGCATCAGCATGTTTATTCTTACTTAACTTC 3131

Qy        781 ATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3132 ATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAAGAGATTGCGAT 3191

Qy        841 AGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3192 AGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCTATCTATATCCT 3251

Qy        901 AAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAATGCGGAAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3252 AAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTTTAATGCGGAAA 3311

Qy        961 TAA 963
              |||
Db       3312 TAA 3314
```

RESULT 2
US-09-816-028A-26
; Sequence 26, Application US/09816028A
; Patent No. US20020042369A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(906)
; OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
; OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
; OTHER INFORMATION: biosynthesis locus)
US-09-816-028A-26

Query Match 5.3%; Score 50.6; DB 3; Length 906;
Best Local Similarity 49.1%; Pred. No. 1.1;
Matches 134; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |          | |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |   |     |  || || |  | || ||||       ||| |   | | |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |   |||||||   | |    |||| | |    ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359
```

RESULT 3
US-10-303-161-26
; Sequence 26, Application US/10303161
; Publication No. US20030148459A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,161
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(906)
; OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
; OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
; OTHER INFORMATION: biosynthesis locus)
US-10-303-161-26

Query Match 5.3%; Score 50.6; DB 6; Length 906;
Best Local Similarity 49.1%; Pred. No. 1.1;
Matches 134; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146
```

```
Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |     |        |  || || |  |  || ||||       ||| |   | |  |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |    |||||||    | |    ||||| |  |    ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 4
US-10-303-118-26
; Sequence 26, Application US/10303118
; Publication No. US20030157655A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,118
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-303-118-26

  Query Match             5.3%;  Score 50.6;  DB 6;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 1.1;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |     |        |  || || |  |  || ||||       ||| |   | |  |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |    |||||||    | |    ||||| |  |    ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 5
US-10-303-128-26
; Sequence 26, Application US/10303128
; Publication No. US20030157656A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,128
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-303-128-26

  Query Match             5.3%;  Score 50.6;  DB 6;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 1.1;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |     |        |  || || |  |  || ||||       ||| |   | |  |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |    |||||||    | |    ||||| |  |    ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 6
US-10-303-134-26
; Sequence 26, Application US/10303134
; Publication No. US20030157657A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,134
```

```
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-303-134-26

  Query Match             5.3%;  Score 50.6;  DB 6;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 1.1;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   ||||  ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |        |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |    |         |  || || |  |  || ||||       ||| |   | |  |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |   |||||||    | |    |||| | |   ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 7
US-10-303-162-26
; Sequence 26, Application US/10303162
; Publication No. US20030157658A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/303,162
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-303-162-26

  Query Match             5.3%;  Score 50.6;  DB 6;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 1.1;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   ||||  ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |        |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |    |         |  || || |  |  || ||||       ||| |   | |  |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |   |||||||    | |    |||| | |   ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 8
US-10-735-419-26
; Sequence 26, Application US/10735419
; Publication No. US20040180406A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/735,419
;  CURRENT FILING DATE:  2003-12-11
;  PRIOR APPLICATION NUMBER: US/09/816,028A
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-735-419-26

  Query Match             5.3%;  Score 50.6;  DB 8;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 1.1;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   ||||  ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
```

```
                ||| |  ||  | |||  | ||  |  |        |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |     |         |  || || |  | || ||||       ||| |  | |  |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |   |||||||    | |   |||||  |  |    ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 9
US-10-820-536-26
; Sequence 26, Application US/10820536
; Publication No. US20040203103A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/820,536
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: 10/303,128
;  PRIOR FILING DATE: 2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-820-536-26

  Query Match             5.3%;  Score 50.6;  DB 8;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 1.1;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||      ||   | |||   | ||  |  |        |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |     |         |  || || |  | || ||||       ||| |  | |  |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |   |||||||    | |   |||||  |  |    ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 10
US-10-845-408-26
; Sequence 26, Application US/10845408
; Publication No. US20040203112A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/845,408
;  CURRENT FILING DATE:  2004-05-12
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-845-408-26

  Query Match             5.3%;  Score 50.6;  DB 8;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 1.1;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||      ||   | |||   | ||  |  |        |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |     |         |  || || |  | || ||||       ||| |  | |  |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |   |||||||    | |   |||||  |  |    ||  | |    || ||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 11
US-10-845-412-26
; Sequence 26, Application US/10845412
; Publication No. US20040203113A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/845,412
```

```
;  CURRENT FILING DATE:  2004-05-12
;  PRIOR APPLICATION NUMBER: US/10/303,128
;  PRIOR FILING DATE: 2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-845-412-26

  Query Match             5.3%;  Score 50.6;  DB 8;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 1.1;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |     |       |  || || |  |  || ||||       ||| |   | |  |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |   |||||||    | |   |||||  |  |    ||  | |    || |||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 12
US-10-846-219-26
; Sequence 26, Application US/10846219
; Publication No. US20040219638A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/846,219
;  CURRENT FILING DATE:  2004-05-14
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-846-219-26

  Query Match             5.3%;  Score 50.6;  DB 8;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 1.1;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
              | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db         87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy        165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
              |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db        147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy        225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
              |     |       |  || || |  |  || ||||       ||| |   | |  |
Db        207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy        285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
              |   |||||||    | |   |||||  |  |    ||  | |    || |||| |
Db        267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy        345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
              | |  || | | ||| | |||| | | ||| |
Db        327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 13
US-10-821-604-26
; Sequence 26, Application US/10821604
; Publication No. US20040229263A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/821,604
;  CURRENT FILING DATE:  2004-04-08
;  PRIOR APPLICATION NUMBER: 10/303,128
;  PRIOR FILING DATE: 2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-821-604-26

  Query Match             5.3%;  Score 50.6;  DB 8;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 1.1;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
```

```
                | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db          87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy         165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
               |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db         147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy         225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
                 |      |         |  || || |  |  || ||||       ||| |   | |  |
Db         207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy         285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
                |    |||||||    | |    ||||| |  |    ||  | |    || ||| |
Db         267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy         345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
               | |  || | | ||| | |||| | | ||| |
Db         327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 14
US-10-847-983-26
; Sequence 26, Application US/10847983
; Publication No. US20040229272A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/847,983
;  CURRENT FILING DATE:  2004-05-17
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-847-983-26

  Query Match             5.3%;  Score 50.6;  DB 8;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 1.1;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy         105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
               | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db          87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy         165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
               |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db         147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy         225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
                 |      |         |  || || |  |  || ||||       ||| |   | |  |
Db         207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy         285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
                |    |||||||    | |    ||||| |  |    ||  | |    || ||| |
Db         267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy         345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
               | |  || | | ||| | |||| | | ||| |
Db         327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


RESULT 15
US-10-821-573-26
; Sequence 26, Application US/10821573
; Publication No. US20040229313A1
; GENERAL INFORMATION:
;  APPLICANT: Gilbert, Michel
;  APPLICANT:  Wakarchuk, Warren W.
;  APPLICANT:  National Research Council of Canada
;  TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
;  TITLE OF INVENTION:  Gangliosides and Ganglioside Mimics
;  FILE REFERENCE: 019633-000111US
;  CURRENT APPLICATION NUMBER: US/10/821,573
;  CURRENT FILING DATE:  2004-04-08
;  PRIOR APPLICATION NUMBER: 10/303,128
;  PRIOR FILING DATE: 2002-11-21
;  PRIOR APPLICATION NUMBER: US/09/816,028
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/118,213
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: US 09/495,406
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 906
;   TYPE: DNA
;   ORGANISM: Campylobacter jejuni
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(906)
;   OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
;   OTHER INFORMATION: OH4384 (ORF 6a of lipooligosaccharide (LOS)
;   OTHER INFORMATION: biosynthesis locus)
US-10-821-573-26

  Query Match             5.3%;  Score 50.6;  DB 8;  Length 906;
  Best Local Similarity   49.1%;  Pred. No. 1.1;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy         105 TGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAGC 164
               | | ||  | ||  | |||||   |||| ||  |  ||  ||| | |    ||  |||||
Db          87 TAAAGATATAGAAATAATTGTAGTTGATGATTGTGGAAATGATAATAGTATAAATATAGC 146

Qy         165 CAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAGG 224
               |||       ||   | |||   | ||  |  |         |  |||||| ||||||||||
Db         147 CAAAGAATACTCTAAAAAAGACAAAAGAATAAAAATAATCCACAATGAAAAAAACTTAGG 206

Qy         225 AATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGAT 284
                 |      |         |  || || |  |  || ||||       ||| |   | |  |
Db         207 TCTTTTAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAACTCTCCTTATATAATGTTTTT 266

Qy         285 TGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGTC 344
                |    |||||||    | |    ||||| |  |    ||  | |    || ||| |
Db         267 AGATCCTGATGATTATTTGGAACTAAATGCTTGTGAAGAGTGTATAAAAATTTTAGATGA 326

Qy         345 CCGGCCTGCATTAGATATGGTGTATGTAAATTC 377
               | |  || | | ||| | |||| | | ||| |
Db         327 ACAGGATGAAGTTGATTTAGTGTTTTTCAATGC 359


Search completed: May 12, 2006, 05:33:50
Job time : 618.595 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:52:36 ; Search time 490.648 Seconds
(without alignments)
8004.956 Million cell updates/sec

Title: US-09-423-093B-3_COPY_2352_3314
Perfect score: 963
Sequence: 1 atgagtaaaaaacctcttct.....cggttttaatgcggaaataa 963

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/2/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/2/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	50.8	5.3	966	18	US-11-079-463-1999	Sequence 1999, Ap
c 2	48.6	5.0	633	7	US-09-925-065A-464691	Sequence 464691,
c 3	48.6	5.0	641	12	US-10-301-480-520881	Sequence 520881,
c 4	48.6	5.0	641	12	US-10-301-480-1134290	Sequence 1134290,
5	48.6	5.0	930	18	US-11-079-463-1998	Sequence 1998, Ap
c 6	47	4.9	609	12	US-10-301-480-537141	Sequence 537141,
c 7	47	4.9	609	12	US-10-301-480-1150550	Sequence 1150550,
c 8	46.4	4.8	586	7	US-09-925-065A-188821	Sequence 188821,
c 9	46.4	4.8	597	12	US-10-301-480-278711	Sequence 278711,
c 10	46.4	4.8	597	12	US-10-301-480-892120	Sequence 892120,
c 11	45.8	4.8	569	7	US-09-925-065A-298888	Sequence 298888,
c 12	45.8	4.8	571	12	US-10-301-480-375511	Sequence 375511,
c 13	45.8	4.8	571	12	US-10-301-480-988920	Sequence 988920,
c 14	45.8	4.8	820	7	US-09-925-065A-6030	Sequence 6030, Ap
c 15	45.8	4.8	820	11	US-10-301-480-107267	Sequence 107267,
c 16	45.8	4.8	820	12	US-10-301-480-720676	Sequence 720676,
c 17	45.6	4.7	674	12	US-10-301-480-541953	Sequence 541953,
c 18	45.6	4.7	674	12	US-10-301-480-1155362	Sequence 1155362,
19	45.6	4.7	913	7	US-09-925-065A-701638	Sequence 701638,
20	45.4	4.7	606	7	US-09-925-065A-404953	Sequence 404953,
c 21	45.4	4.7	394468	10	US-10-995-561-13473	Sequence 13473, A
c 22	45.2	4.7	586	7	US-09-925-065A-188819	Sequence 188819,
c 23	45.2	4.7	586	7	US-09-925-065A-188820	Sequence 188820,
c 24	45.2	4.7	586	7	US-09-925-065A-188822	Sequence 188822,
c 25	45.2	4.7	597	12	US-10-301-480-278709	Sequence 278709,
c 26	45.2	4.7	597	12	US-10-301-480-278710	Sequence 278710,
c 27	45.2	4.7	597	12	US-10-301-480-278712	Sequence 278712,
c 28	45.2	4.7	597	12	US-10-301-480-604712	Sequence 604712,
c 29	45.2	4.7	597	12	US-10-301-480-892118	Sequence 892118,
c 30	45.2	4.7	597	12	US-10-301-480-892119	Sequence 892119,
c 31	45.2	4.7	597	12	US-10-301-480-892121	Sequence 892121,
c 32	45.2	4.7	597	12	US-10-301-480-1218121	Sequence 1218121,
c 33	45	4.7	600	17	US-11-136-527-7542	Sequence 7542, Ap
c 34	45	4.7	636	7	US-09-925-065A-510391	Sequence 510391,
c 35	45	4.7	636	7	US-09-925-065A-510392	Sequence 510392,
c 36	45	4.7	6000	17	US-11-136-527-3446	Sequence 3446, Ap
c 37	44.8	4.7	1342	7	US-09-925-065A-703795	Sequence 703795,
c 38	44.8	4.7	2068	7	US-09-925-065A-670498	Sequence 670498,
39	44.8	4.7	9347	10	US-10-240-708-36	Sequence 36, Appl
40	44.8	4.7	225587	11	US-10-330-773-374	Sequence 374, App
c 41	44.2	4.6	987	12	US-10-301-480-545289	Sequence 545289,
c 42	44.2	4.6	987	12	US-10-301-480-1158698	Sequence 1158698,
43	44.2	4.6	173602	17	US-11-121-086-25	Sequence 25, Appl
c 44	43.8	4.5	525	7	US-09-925-065A-192012	Sequence 192012,
c 45	43.8	4.5	528	12	US-10-301-480-281590	Sequence 281590,

ALIGNMENTS

RESULT 1
US-11-079-463-1999
; Sequence 1999, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 1999
; LENGTH: 966
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-1999

Query Match 5.3%; Score 50.8; DB 18; Length 966;
Best Local Similarity 47.9%; Pred. No. 3.7;
Matches 182; Conservative 0; Mismatches 192; Indels 6; Gaps 1;

```
Qy          10 AAACCTCTTCTTACTATTGCTATTCCGACATATAACCGCTCTTCATGTTTGGCTCGTTTA 69
               ||   | |  |  | |||  ||| || || |||||  | | |     ||| | |  | |
Db         124 AATAATTTGTTGTCCATTTTTATACCTACTTATAATAGGTGTAATGTTTTAGATACTGTT 183

Qy          70 CTTGATAGTATAATTCAACAGGAGAACTATTGTCATGATGAACTCGAGGTTATTGTTTGT 129
               | ||| || | |||       |  ||  | |       ||||  |    ||    || | |
Db         184 TTAGATCGTGTCATTAGTTCTGTAAAAGAAT------ATGATGTTTGTGTACAAGTGTAT 237

Qy         130 GATAATGCTTCAACAGATGAAACAGCAAGAATAGCCAAGAGTGGCTTAGATAAAATAAGA 189
               ||||||  |||    ||| |||||| ||  || |             | |
Db         238 GATAATTGTTCTCTTGATAAAACAGAAAATATTGTATGTGCAAAGCAAAAGTTGTATCCT 297
```

```
Qy        190 AATAGTACTTATCATCTAAATGAAGAAAACTTAGGAATGGATGGTAACTTCCAGAAATGT 249
              |||      |||  |  |||||||   ||  ||||  | ||| | ||  | |  ||
Db        298 TATATATTCTATTGTAAAAATGAAACTAATATAGGGCTAGATCGAAATATGCTTAATGTC 357

Qy        250 TTTGAGTTATCAAATGGAAAATATCTTTGGATGATTGGCGATGATGATCTAATAGTCAAA 309
              || |   |  |      ||||||   |  ||| | || |||||||||  |  | ||  |
Db        358 CTTAATAAACAAGTCTCTAAATATACATTAATGTTGGGAGATGATGATTTTTTGGTAGAT 417

Qy        310 AATGGTATTTCGAAGGTTTTTTCGATATTAAAGTCCCGGCCTGCATTAGATATGGTGTAT 369
              ||  |  |  |||  |    ||    |||||     |  ||  ||   | |  | | |
Db        418 GATTTTTATAAGAAAATAGAATCTTACTTAAAACAATGTTTTGATTTTATTGTCCTTTCT 477

Qy        370 GTAAATTCAGCAGCAAAGAC 389
              ||  ||||  ||   || ||
Db        478 GTTGATTCGACATTTAAAAC 497


RESULT 2
US-09-925-065A-464691/c
; Sequence 464691, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 464691
;   LENGTH: 633
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-464691

  Query Match             5.0%;  Score 48.6;  DB 7;  Length 633;
  Best Local Similarity   54.9%;  Pred. No. 8.8;
  Matches   96;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        764 TATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATT 823
              |||  |    | |   | |||  | |    | || | |   |||  | || || |||||  ||
Db        406 TATATTATATAATTATGATATACAAAAGATATGATATGCAAAGATATGATAAAAATATTT 347

Qy        824 ATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGT 883
              ||||  | |  |||  |   |  ||  ||||||  | ||||| ||  | ||   | ||  |
Db        346 ATTTTAAAATTTGTAAGTATTTATCAATAGATACATTTATAAATTTAAATATTCATATTT 287

Qy        884 TTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTAT 938
              ||  ||  |||||  ||||  |      |    |||  ||| |||   || ||||
Db        286 ATTATTAAATATATGATAAACTTAAAAATAATATTAATAAACTAATTAATTTTAT 232


RESULT 3
US-10-301-480-520881/c
; Sequence 520881, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 520881
;   LENGTH: 641
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-520881

  Query Match             5.0%;  Score 48.6;  DB 12;  Length 641;
  Best Local Similarity   54.9%;  Pred. No. 8.9;
  Matches   96;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        764 TATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATT 823
              |||  |    | |   | |||  | |    | || | |   |||  | || || |||||  ||
Db        406 TATATTATATAATTATGATATACAAAAGATATGATATGCAAAGATATGATAAAAATATTT 347

Qy        824 ATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGT 883
              ||||  | |  |||  |   |  ||  ||||||  | ||||| ||  | ||   | ||  |
Db        346 ATTTTAAAATTTGTAAGTATTTATCAATAGATACATTTATAAATTTAAATATTCATATTT 287

Qy        884 TTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTAT 938
              ||  ||  |||||  ||||  |      |    |||  ||| |||   || ||||
Db        286 ATTATTAAATATATGATAAACTTAAAAATAATATTAATAAACTAATTAATTTTAT 232


RESULT 4
US-10-301-480-1134290/c
; Sequence 1134290, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1134290
;   LENGTH: 641
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1134290

  Query Match             5.0%;  Score 48.6;  DB 12;  Length 641;
  Best Local Similarity   54.9%;  Pred. No. 8.9;
  Matches   96;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        764 TATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATT 823
              |||  |    | |   | |||  | |    | || | |   |||  | || || |||||  ||
Db        406 TATATTATATAATTATGATATACAAAAGATATGATATGCAAAGATATGATAAAAATATTT 347

Qy        824 ATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGT 883
              ||||  | |  |||  |   |  ||  ||||||  | ||||| ||  | ||   | ||  |
Db        346 ATTTTAAAATTTGTAAGTATTTATCAATAGATACATTTATAAATTTAAATATTCATATTT 287

Qy        884 TTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTAT 938
              ||  ||  |||||  ||||  |      |    |||  ||| |||   || ||||
Db        286 ATTATTAAATATATGATAAACTTAAAAATAATATTAATAAACTAATTAATTTTAT 232


RESULT 5
US-11-079-463-1998
```

```
; Sequence 1998, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PATH00-03DIV2
;  CURRENT APPLICATION NUMBER: US/11/079,463
;  CURRENT FILING DATE:  2005-03-14
;  PRIOR APPLICATION NUMBER: US 60/128,705
;  PRIOR FILING DATE: 1999-04-09
;  PRIOR APPLICATION NUMBER: US 09/540,209
;  PRIOR FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 1998
;   LENGTH: 930
;   TYPE: DNA
;   ORGANISM: B.fragilis
US-11-079-463-1998

  Query Match             5.0%;  Score 48.6;  DB 18;  Length 930;
  Best Local Similarity   46.4%;  Pred. No. 9.4;
  Matches  159;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy        104 ATGATGAACTCGAGGTTATTGTTTGTGATAATGCTTCAACAGATGAAACAGCAAGAATAG 163
              || |  |  | ||| ||||| |   ||||||   |  ||||||  |||    || ||
Db        107 ATCAAAATATTGAGATTATTATAGTTGATAACTGTAGTACAGATTTAACTTGGAGTATTT 166

Qy        164 CCAAGAGTGGCTTAGATAAAATAAGAAATAGTACTTATCATCTAAATGAAGAAAACTTAG 223
                || |   |   |   |||      |  | ||      |||  ||| ||   ||  |||
Db        167 TGAATACATGGGCAAGCAAAGATCATAGAATTAAAATATATCAGAATAAAACTAATATAG 226

Qy        224 GAATGGATGGTAACTTCCAGAAATGTTTTGAGTTATCAAATGGAAAATATCTTTGGATGA 283
              ||   | |   || |   |  |||| ||| ||    |   ||  | ||| |   |||
Db        227 GACCTGTTTTAAATTGGAATGAATGCTTTAAGCATGCTTCAGGTGAGTATATAAAGATTC 286

Qy        284 TTGGCGATGATGATCTAATAGTCAAAAATGGTATTTCGAAGGTTTTTTCGATATTAAAGT 343
              || |   |||||||   || |    | ||  ||||   |  |   |     || |  |
Db        287 TTTGGTCTGATGATTGGATGGCATTAGATTTTATTGAAAGAGCAGTGAATCTAATTGATG 346

Qy        344 CCCGGCCTGCATTAGATATGGTGTATGTAAATTCAGCAGCAAAGACTGAGTTAAACTATA 403
                | | ||  ||  | |||      ||   ||    | | || | | ||   |     |||
Db        347 AGAAGTCAGCGTTTGTTATATCAAATCATAAGATCGTATCAGAAAATGGTATTGTAGATA 406

Qy        404 ATGCTGATGTGAGGACGTCATTCTACACAAATGATGTAGATTT 446
              |||   |    | ||    |   || ||||   | |   || |
Db        407 ATGTAAAGTATAAGAAACAAAAATATACAAGAAAAGAGTATCT 449


RESULT 6
US-10-301-480-537141/c
; Sequence 537141, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 537141
;   LENGTH: 609
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 251
;   OTHER INFORMATION: n = A,T,C or G
US-10-301-480-537141

  Query Match             4.9%;  Score 47;  DB 12;  Length 609;
  Best Local Similarity   60.6%;  Pred. No. 17;
  Matches   77;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        817 AATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGG 876
              |||| |||| ||  |  ||     |||  || | |  || || || ||||||| |||
Db        566 AATATTTATATATAATTTTATATAAGTATATATTTCTATATATTTTTATATAAGTATATA 507

Qy        877 CTTAGGTTTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATT 936
              |||  | | | ||  |||||  | | | | |   | ||||||  | | |||| || |||
Db        506 TTTATATATATATAAATATATATTTATAATATTTATATATTTATATATATAATATTTATA 447

Qy        937 ATAAAGA 943
              |||  | |
Db        446 ATATAAA 440


RESULT 7
US-10-301-480-1150550/c
; Sequence 1150550, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1150550
;   LENGTH: 609
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 251
;   OTHER INFORMATION: n = A,T,C or G
US-10-301-480-1150550

  Query Match             4.9%;  Score 47;  DB 12;  Length 609;
  Best Local Similarity   60.6%;  Pred. No. 17;
  Matches   77;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        817 AATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGG 876
              |||| |||| ||  |  ||     |||  || | |  || || || ||||||| |||
Db        566 AATATTTATATATAATTTTATATAAGTATATATTTCTATATATTTTTATATAAGTATATA 507

Qy        877 CTTAGGTTTTTCTATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATT 936
              |||  | | | ||  |||||  | | | | |   | ||||||  | | |||| || |||
Db        506 TTTATATATATATAAATATATATTTATAATATTTATATATTTATATATATAATATTTATA 447

Qy        937 ATAAAGA 943
              |||  | |
Db        446 ATATAAA 440


RESULT 8
US-09-925-065A-188821/c
; Sequence 188821, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
```

```
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 188821
;   LENGTH: 586
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-188821

  Query Match             4.8%;  Score 46.4;  DB 7;  Length 586;
  Best Local Similarity   52.1%;  Pred. No. 22;
  Matches  101;  Conservative    1;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        770 TACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAA 829
              | | |||| |   || | ||             ||| || ||  || | |||  ||| |||
Db        461 TTCATAACCTGTGAGACAATTCCTTATTTTTAAAATATATTTATATATATATATATATAA 402

Qy        830 GAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCT 889
              | |||   |||       | ||| || ||| |||||||| | ||     ||   | |
Db        401 TATATTTATATATAAATATAATATATATAAATATATATATAAATATATATATAATATATA 342

Qy        890 ATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTT 949
              :          ||||  | |   | ||| || | | ||||||||| |||| | |   | |
Db        341 WAAATATATATAAATTTATATTTATATATAAATATATAAAATATAATATATATAATATAT 282

Qy        950 TAATGCGGAAATAA 963
                ||    | ||||
Db        281 ATATTTATATATAA 268


RESULT 9
US-10-301-480-278711/c
; Sequence 278711, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 278711
;   LENGTH: 597
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-278711

  Query Match             4.8%;  Score 46.4;  DB 12;  Length 597;
  Best Local Similarity   52.1%;  Pred. No. 22;
  Matches  101;  Conservative    1;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        770 TACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAA 829
              | | |||| |   || | ||             ||| || ||  || | |||  ||| |||
Db        461 TTCATAACCTGTGAGACAATTCCTTATTTTTAAAATATATTTATATATATATATATATAA 402

Qy        830 GAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCT 889
              | |||   |||       | ||| || ||| |||||||| | ||     ||   | |
Db        401 TATATTTATATATAAATATAATATATATAAATATATATATAAATATATATATAATATATA 342

Qy        890 ATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTT 949
              :          ||||  | |   | ||| || | | ||||||||| |||| | |   | |
Db        341 WAAATATATATAAATTTATATTTATATATAAATATATAAAATATAATATATATAATATAT 282

Qy        950 TAATGCGGAAATAA 963
                ||    | ||||
Db        281 ATATTTATATATAA 268


RESULT 10
US-10-301-480-892120/c
; Sequence 892120, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 892120
;   LENGTH: 597
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-892120

  Query Match             4.8%;  Score 46.4;  DB 12;  Length 597;
  Best Local Similarity   52.1%;  Pred. No. 22;
  Matches  101;  Conservative    1;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        770 TACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGCTACAAATAATTATTTAA 829
              | | |||| |   || | ||             ||| || ||  || | |||  ||| |||
Db        461 TTCATAACCTGTGAGACAATTCCTTATTTTTAAAATATATTTATATATATATATATATAA 402

Qy        830 GAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATATGGGCTTAGGTTTTTCT 889
              | |||   |||       | ||| || ||| |||||||| | ||     ||   | |
Db        401 TATATTTATATATAAATATAATATATATAAATATATATATAAATATATATATAATATATA 342

Qy        890 ATCTATATCCTAAAACTGTGCCTTTATTTAGAAAAATAAAATATATTATAAAGACGGTTT 949
              :          ||||  | |   | ||| || | | ||||||||| |||| | |   | |
Db        341 WAAATATATATAAATTTATATTTATATATAAATATATAAAATATAATATATATAATATAT 282

Qy        950 TAATGCGGAAATAA 963
                ||    | ||||
Db        281 ATATTTATATATAA 268


RESULT 11
US-09-925-065A-298888/c
; Sequence 298888, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
```

```
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 298888
;   LENGTH: 569
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-298888

  Query Match             4.8%;  Score 45.8;  DB 7;  Length 569;
  Best Local Similarity   61.2%;  Pred. No. 29;
  Matches   74;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        753 ATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGC 812
              || |  ||  || || | | | || ||||  | | |||| || || ||  |||||| |
Db        381 ATAAAAATTATTTTTTTAATTAAAATTCAGTGCCAATGAGGAAAATACTTAAAATTCTAA 322

Qy        813 TACAAATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATA 872
              || |  |||   ||  |  ||   |  || ||| |||| || ||  |||||||| |||
Db        321 ACCATAGTATTGAATATAATTTTAGAAATGTACATATATATATATATATATATATATATA 262

Qy        873 T 873
              |
Db        261 T 261


RESULT 12
US-10-301-480-375511/c
; Sequence 375511, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 375511
;   LENGTH: 571
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-375511

  Query Match             4.8%;  Score 45.8;  DB 12;  Length 571;
  Best Local Similarity   61.2%;  Pred. No. 29;
  Matches   74;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        753 ATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTGC 812
              || |  ||  || || | | | || ||||  | | |||| || || ||  |||||| |
Db        383 ATAAAAATTATTTTTTTAATTAAAATTCAGTGCCAATGAGGAAAATACTTAAAATTCTAA 324

Qy        813 TACAAATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATA 872
              || |  |||   ||  |  ||   |  || ||| |||| || ||  |||||||| |||
Db        323 ACCATAGTATTGAATATAATTTTAGAAATGTACATATATATATATATATATATATATATA 264

Qy        873 T 873
              |
Db        263 T 263


RESULT 13
US-10-301-480-988920/c
; Sequence 988920, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 988920
;   LENGTH: 571
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-988920

  Query Match             4.8%;  Score 45.8;  DB 12;  Length 571;
  Best Local Similarity   61.2%;  Pred. No. 29;
  Matches   74;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        753 ATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTGC 812
              || |  ||  || || | | | || ||||  | | |||| || || ||  |||||| |
Db        383 ATAAAAATTATTTTTTTAATTAAAATTCAGTGCCAATGAGGAAAATACTTAAAATTCTAA 324

Qy        813 TACAAATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATA 872
              || |  |||   ||  |  ||   |  || ||| |||| || ||  |||||||| |||
Db        323 ACCATAGTATTGAATATAATTTTAGAAATGTACATATATATATATATATATATATATATA 264

Qy        873 T 873
              |
Db        263 T 263


RESULT 14
US-09-925-065A-6030/c
; Sequence 6030, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6030
;   LENGTH: 820
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-6030

  Query Match             4.8%;  Score 45.8;  DB 7;  Length 820;
  Best Local Similarity   61.2%;  Pred. No. 30;
  Matches   74;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        753 ATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTGC 812
              || |  ||  || || | | | || ||||  | | |||| || || ||  |||||| |
Db        681 ATAAAAATTATTTTTTTAATTAAAATTCAGTGCCAATGAGGAAAATACTTAAAATTCTAA 622
```

```
Qy        813 TACAAATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATA 872
              || |  |||   ||  |  ||   |  || ||| |||| || ||  |||||||| |||
Db        621 ACCATAGTATTGAATATAATTTTAGAAATGTACATATATATATATATATATATATATATA 562

Qy        873 T 873
              |
Db        561 T 561


RESULT 15
US-10-301-480-107267/c
; Sequence 107267, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 107267
;   LENGTH: 820
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-107267

  Query Match             4.8%;  Score 45.8;  DB 11;  Length 820;
  Best Local Similarity   61.2%;  Pred. No. 30;
  Matches   74;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        753 ATCAGCATGTTTATTCTTACTTAACTTCATAGGCGATGAAGATAAAACCAAAAATTTTGC 812
              || |  ||  || || | | | || ||||  | | |||| || || ||  |||||| |
Db        681 ATAAAAATTATTTTTTTAATTAAAATTCAGTGCCAATGAGGAAAATACTTAAAATTCTAA 622

Qy        813 TACAAATAATTATTTAAGAGATTGCGATAGTGCATTTATAGATTTAATTATATATAAATA 872
              || |  |||   ||  |  ||   |  || ||| |||| || ||  |||||||| |||
Db        621 ACCATAGTATTGAATATAATTTTAGAAATGTACATATATATATATATATATATATATATA 562

Qy        873 T 873
              |
Db        561 T 561


Search completed: May 12, 2006, 06:54:53
Job time : 491.648 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 08:31:04 ; Search time 1853.66 Seconds
(without alignments)
15792.702 Million cell updates/sec

Title: US-09-423-093B-3_COPY_3361_3875
Perfect score: 515
Sequence: 1 atgtttgtctatagtttaag.....gctggtagtgaaatttgtga 515

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	515	100.0	12441	1	SERFBC2	X61917 S.enterica
	2	515	100.0	12441	6	BD085556	BD085556 Nucleic a
	3	515	100.0	12446	6	AX029263	AX029263 Sequence
	4	59.4	11.5	24454	1	AF048749	AF048749 Bacteroid
	5	59.4	11.5	110000	1	CR626927_12	Continuation (13 o
	6	57.8	11.2	110000	1	CP000076_58	Continuation (59 o
c	7	54.6	10.6	199133	5	BX323062	BX323062 Zebrafish
	8	53.2	10.3	349980	6	AX344563	AX344563 Sequence
	9	53	10.3	1589	2	AF275222	AF275222 Schlechte
	10	52	10.1	110000	14	PFMAL13_18	Continuation (19 o
	11	52	10.1	320170	14	CR792440	CR792440 Danio rer
c	12	51.2	9.9	184622	14	CR848702	CR848702 Danio rer
	13	50.8	9.9	14263	1	AF172324	AF172324 Escherich
c	14	50.8	9.9	219074	14	CR769779	CR769779 Danio rer
	15	50.4	9.8	228279	14	CR848045	CR848045 Danio rer
c	16	50.4	9.8	228327	14	CR847824	CR847824 Danio rer
	17	50.2	9.7	187186	14	CR392353	CR392353 Danio rer
c	18	49.8	9.7	3188	2	AY601893	AY601893 Schlechte
c	19	49.6	9.6	408	6	AX300942	AX300942 Sequence
c	20	49.6	9.6	103700	5	BX510946	BX510946 Zebrafish
c	21	49.6	9.6	169361	5	CR762498	CR762498 Zebrafish
	22	49	9.5	1900	2	MELMTCOII	L76262 Meloidogyne
c	23	49	9.5	85779	15	SCE011856	AJ011856 Saccharom
c	24	49	9.5	110000	14	BX890596_1	Continuation (2 of
	25	49	9.5	114579	9	AC084072	AC084072 Mus muscu
	26	49	9.5	125958	2	AC115592	AC115592 Dictyoste
c	27	49	9.5	182223	14	CR790384	CR790384 Danio rer
	28	49	9.5	246587	9	BX072536	BX072536 Mouse DNA
	29	48.8	9.5	8711	6	AX251456	AX251456 Sequence
	30	48.8	9.5	21277	1	AF343089	AF343089 Aeromonas
	31	48.8	9.5	21279	1	AF148126	AF148126 Aeromonas
c	32	48.4	9.4	110000	1	BA000021_3	Continuation (4 of
c	33	48.4	9.4	188314	14	CR774179	CR774179 Danio rer
c	34	48	9.3	252704	14	CR792433	CR792433 Danio rer
	35	47.6	9.2	11422	6	AX345120	AX345120 Sequence
	36	47.6	9.2	11422	6	AX348322	AX348322 Sequence
	37	47.6	9.2	196555	14	CR318605	CR318605 Danio rer
c	38	47.6	9.2	231293	14	CR318596	CR318596 Danio rer
c	39	47.4	9.2	517	15	MISCTRLP	X06424 Yeast mit D
	40	47.4	9.2	162952	14	CR388068	CR388068 Danio rer
c	41	47.4	9.2	174427	14	CR812798	CR812798 Danio rer
	42	47.2	9.2	17967	6	AX345916	AX345916 Sequence
c	43	47.2	9.2	236005	14	CR627492	CR627492 Danio rer
c	44	47	9.1	110000	14	AL583837_1	Continuation (2 of
c	45	47	9.1	110000	15	AP008216_092	Continuation (93 o

ALIGNMENTS

RESULT 1
SERFBC2
LOCUS SERFBC2 12441 bp DNA linear BCT 09-SEP-2004
DEFINITION S.enterica rfbJ gene cluster.
ACCESSION X61917
VERSION X61917.1 GI:47004
KEYWORDS abe gene; abequose synthase; galactosyltransferase; manB gene; manC gene; mannose-1-phosphate guanyltransferase; phosphomannomutase; wbaL gene; wbaP gene; wbaQ gene; wbaR gene; wbaW gene; wbaZ gene; wzx gene; wzy gene.
SOURCE Salmonella enterica
ORGANISM Salmonella enterica
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 12441)
AUTHORS Brown,P.K., Romana,L.K. and Reeves,P.R.
TITLE Molecular analysis of the rfb gene cluster of Salmonella serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B
JOURNAL Mol. Microbiol. 6 (10), 1385-1394 (1992)
PUBMED 1379320
REFERENCE 2
AUTHORS Reeves,P.
TITLE Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale
JOURNAL Trends Genet. 9 (1), 17-22 (1993)
PUBMED 8434412
REFERENCE 3
AUTHORS Liu,D., Haase,A.M., Lindqvist,L., Lindberg,A.A. and Reeves,P.R.
TITLE Glycosyl transferases of O-antigen biosynthesis in Salmonella enterica: identification and characterization of transferase genes of groups B, C2, and E1
JOURNAL J. Bacteriol. 175 (11), 3408-3413 (1993)
PUBMED 7684736
REFERENCE 4 (bases 1 to 12441)
AUTHORS Brown,P.K.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1991) P.K. Brown, University of Sydney, Dept of Microbiology, Maze Crescent, Sydney N S W 2006, AUSTRALIA
FEATURES Location/Qualifiers
source 1. .12441

/organism="Salmonella enterica"
/mol_type="genomic DNA"
/strain="group C2, serovar muenchen M67"
/db_xref="taxon:28901"

gene 138. .1019
/gene="abe"

CDS 138. .1019
/gene="abe"
/codon_start=1
/transl_table=11
/product="abequose synthase"
/protein_id="CAA43918.1"
/db_xref="GI:47005"
/db_xref="GOA:Q00329"
/db_xref="UniProt/Swiss-Prot:Q00329"
/translation="MLDVNKKILMTGATSFVGTHLLHSLIKEGYSIIALKRPITEPTI
INTLIEWLNIQDIEKICQSSMNIHAIVHIATDYGRNRTPISEQYKCNVLLPTRLLELM
PALKTKFFISTDSFFGKYEKHYGYMRSYMASKRHFVELSKIYVEEHPDVCFINLRLEH
VYGERDKAGKIIPYVIKKMKNNEDIDCTIARQKRDFIYIDDVVSAYLKILKEGFNAGH
YDVEVGTGKSIELKEVFEIIKKETHSSSKINYGAVAMRDDEIMESHANTSFLTRLGWS
AEFSIEKGVKKMLSMKE"

CDS 1019. .2359
/note="wzx
Referred to as orf12.6 in reference [2]"
/citation=[2]
/codon_start=1
/transl_table=11
/product="hypothetical protein with 12 predicted
transmembrane segments which possibly functions in export
of O-antigen"
/protein_id="CAA43919.1"
/db_xref="GI:47006"
/db_xref="UniProt/TrEMBL:Q00339"
/translation="MNRIIRMLGVDKAIRYVIFGKIISVLTGLLLIMLISHHLSKDAQ
GYYYTFNSVVALQIIFELGLSTVIIQFASHEMSALKYDYSERDIIGESKNKQRYLSLF
RLAIKWYAVIALLIILIVGPIGYVFFTQKEGLGVPWQGAWLLLTIVTAFNIFLVSVLS
VAEGSGLITDVNKMRMYQSLLAGILAVSLLISGFGLYATSAIAISGTIIFSIFSYKYF
KKIFLQSLKHKNKYTEGGISWVNEIFPMQWRIALSWMSGYFIYFVMTPIAFKYFGAIY
AGQLGMSLTLCNMVMATGLAWISTKYPKWGVMVSNKQLAELSKSFKSAVMQSSFFVLT
GLTGVYISLWLLKLSGSNIGERFLGLQDFFFLSLAIIGNHIVACFATYIRAHKTEKMT
LASCIMALLTITTMLFVAYLEYSRFYMLMYAALTWLYFVPQTYIIFKRFKSSYE"

CDS 2352. .3314
/note="unnamed protein product; wbaR"
/codon_start=1
/transl_table=11
/protein_id="CAA43909.1"
/db_xref="GI:47007"
/db_xref="UniProt/TrEMBL:Q00337"
/translation="MSKKPLLTIAIPTYNRSSCLARLLDSIIQQENYCHDELEVIVCD
NASTDETARIAKSGLDKIRNSTYHLNEENLGMDGNFQKCFELSNGKYLWMIGDDDLIV
KNGISKVFSILKSRPALDMVYVNSAAKTELNYNADVRTSFYTNDVDFISDVKVMFTFI
SGMICKKTDAIVKAVGIFSPQTTGKYLMHLTWQLPLLKQGGEFAVIHNNIIEAEPDNS
GGYHLYKVFSNNLATIFDVFYPREHRVSKRVRASACLFLLNFIGDEDKTKNFATNNYL
RDCDSAFIDLIIYKYGLRFFYLYPKTVPLFRKIKYIIKTVLMRK"

CDS 3361. .3975
/note="unnamed protein product; wbaL
orf14.9 possibly encodes an acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="CAA43910.1"
/db_xref="GI:47008"
/db_xref="UniProt/TrEMBL:Q00480"
/translation="MFVYSLRLKLNLIISLLSKVRRKSKAKFLVLLSGYDFKMVGKNF
KLNVKPYSAKNNTSSKWGSMRVGDNCWIEAVYNYGDEKFEPYLYIGDRICLSDNVHIS
CVSCLILENDILIGSKVYIGDHSHGSYKVCSPKIEPPANKPLGDIAPIKIGNCCWIGD
NAVILAGSEICDGCVIAANSVVKDLKVDKPCLIGGVPAKVIKVF"

CDS 3977. .5020
/note="wbaQ
Referred to as orf15.6 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="rhamnosyl transferase"
/protein_id="CAA43911.1"
/db_xref="GI:47009"
/db_xref="GOA:Q00482"
/db_xref="UniProt/TrEMBL:Q00482"
/translation="MNVFISICIPSYNRAEFLEPLLDSIYNQDYCLKNNDFEVIVCED
KSPQRDEINSIIENYKAKNNKQNLYVNFNEDNLGYDKNLKKCISLTTGKYCMIMGNDD
LLADGALSKIVKVLKANPEIVLATRAYGWFKENPNELCDTVRHLTDDTLFQPGADAIK
FFFRRVGVISGFIVNAEKAKKLSSDLFDGRLYYQMYLAGMLMAEGQGYYFSDVMTLSR
DTEAPDFGNAGTEKGVFTPGGYKPEGRIHMVEGLLLIAKYIEDTTKIDGVYAGIRKDL
ANYFYPYIRDQLDLPLYTYIKMINKFRKMGFSNEKLFYVHAFLGYVLKRRGYDALIKY
IRSKKGGTPRLGI"

gene 5114. .6313
/gene="wzy"

CDS 5114. .6313
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O antigen polymerase"
/protein_id="CAA43912.1"
/db_xref="GI:47010"
/db_xref="GOA:Q00474"
/db_xref="UniProt/Swiss-Prot:Q00474"
/translation="MLPFPPGAILRDVLNVFFVALVLVRFVIDRKKTYFPLVFTIFSW
SAVILWVIALTIFSPDKIQAIMGGRSYILFPAVFIALVILKVSYPQSLNIEKIVCYII
FLMFMVATISIIDVLMNGEFIKLLGYDEHYAGEQLNLINSYDGMVRATGGFSDALNFG
YMLTLGVLLCMECFSQGYKRLLMLIISFVLFIAICMSLTRGAILVAALIYALYIISNR
KMLFCGITLFVIIPVLAISTNIFDNYTEILIGRFTDSSQASRGSTQGRIDMAINSLN
FLSEHPSGIGLGTQGSGNMLSVKDNRLNTDNYFFWIALETGIIGLIINIIYLASQFYS
STLLNRIYGSHCSNMHYRLYFLFGSIYFISAALSSAPSSSTFSIYYWTVLALIPFLKL
TNRRCTR"

CDS 6313. .7323
/note="wbaW
Referred to as orf17.9 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="second mannosyl transferase"
/protein_id="CAA43913.1"
/db_xref="GI:47011"
/db_xref="GOA:Q00481"
/db_xref="UniProt/TrEMBL:Q00481"
/translation="MNNKKVLMDISWSNKGGIGRFTDEISKLLCDISKEELYRKCASP
LAPLGLAVNIFLRKKTDVVFLPGYIPPLFCSKKFIITIHDLNHLDLNDNSSLFKRLFY
NFIIKRGCRKAYKIFTVSNFSKERIVAWSGVNPNKIVTVYNGVSSLFNADVKPLNLGY
KYLLCVGNRKTHKNEKCVISAFAKADIDPSIKLVFTGNPCNDLEKLIIQHGLSERVKF
FGFVSEKDLPSLYKGSLGLVFPSLYEGFGLPVVEGMACGIPVLTSLTSSLPEVAGDAA
ILVDPLSEDAITKGISRLINDSELRKHLIQKGLLRAKRFNWQNVVSEIEMVLTEACDG
NK"

CDS 7310. .8467
/note="wbaZ
Referred to as orf18.9 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="first mannosyl transferase"
/protein_id="CAA43914.1"
/db_xref="GI:47012"
/db_xref="GOA:Q00483"
/db_xref="UniProt/TrEMBL:Q00483"
/translation="MEINEIKISLVHEWLLSYAGSEQVSSAILHVFPEAKLYSVVDFL
TDEQRRHFLGKYATTTFIQNLPKAKKFYQKYLPLMPLAIEQLDLSDANIIISSAHSVA
KGVISGPDQLHISYVHSPIRYAWDLQHQYLNESNLNKGIKGWLAKWLLHKIRIWDSRT
ANGVDHFIANSQYIARRIKKVYRREASVIYPPVDVDNFEVKNEKQDYYFTASRMVPYK
RIDLIVEAFSKMPEKKLVVIGDGPEMKKIKSKATDNIKLLGYQSFPVLKEYMQSARAF
VFAAEEDFGIIPVEAQACGTPVIAFGKGGALETVRPLGVEEPTGIFFKEQNIASLHEA
VSEFEKNASFFTSQACRKNAEKFSRSRFEQEFKNFVNEKWNLFKTEQIIKR"

gene 8482. .9903
/gene="manC"

CDS 8482. .9903
/gene="manC"
/codon_start=1
/transl_table=11
/product="mannose-1-phosphate guanyltransferase"

```
                     /protein_id="CAA43915.1"
                     /db_xref="GI:47013"
                     /db_xref="GOA:Q00473"
                     /db_xref="UniProt/Swiss-Prot:Q00473"
                     /translation="MSKLIPVIMAGGIGSRLWPLSREEHPKQFLSVDGELSMLQNTIK
                     RLTPLLAGEPLVICNDSHRFLVAEQLRAINKLANNIILEPVGRNTAPAIALAAFCSLQ
                     NVVDEDPLLLVLAADHVIRDEKVFLKAINHAEFFATQGKLVTFGIVPTQAETGYGYIC
                     RGEAIGEDAFSVAEFVEKPDFDTARHYVESEKYYWNSGMFLFRASSYLQELKDLSPDI
                     YQACENAVGSINPDLDFIRIDKEAFAMCPSDSIDYAVMEHTRHAVVVPMNAGWSDVGS
                     WSSLWDISKKDPQRNVLHGDIFAYNSKDNYIYSEKSFISTIGVNNLVIVQTADALLVS
                     DKDSVQDVKKVVDYLKANNRNEHKKHLEVFRPWGKFSVIHSGDNYLVKRITVKPGAKF
                     AAQMHLHRAEHWIVVSGTACITKGEEIFTISENESTFIPANTVHTLKNPATIPLELIE
                     IQSGTYLAEDDIIRLEKHSGYLE"
     gene            9911. .11347
                     /gene="manB"

  Query Match           100.0%;  Score 515;  DB 1;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-83;
  Matches  515;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 3420

Qy         61 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 3480

Qy        121 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 3540

Qy        181 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 3600

Qy        241 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 3660

Qy        301 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT 3720

Qy        361 ATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 ATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA 3780

Qy        421 AATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 AATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGAT 3840

Qy        481 AATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 515
              |||||||||||||||||||||||||||||||||||
Db       3841 AATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 3875


RESULT 2
BD085556
LOCUS       BD085556               12441 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085556
VERSION     BD085556.1  GI:22631166
KEYWORDS    JP 2001522248-A/3.
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 12441)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 3 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Salmonella enterica
            PN   JP 2001522248-A/3
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .12441
            FT                   /organism='Salmonella enterica'.
FEATURES             Location/Qualifiers
     source          1. .12441
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match           100.0%;  Score 515;  DB 6;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-83;
  Matches  515;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 3420

Qy         61 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 3480

Qy        121 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 3540

Qy        181 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 3600

Qy        241 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 3660

Qy        301 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT 3720

Qy        361 ATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 ATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA 3780

Qy        421 AATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 AATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGAT 3840

Qy        481 AATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 515
              |||||||||||||||||||||||||||||||||||
Db       3841 AATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 3875


RESULT 3
AX029263
LOCUS       AX029263               12446 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 3 from Patent WO9850531.
ACCESSION   AX029263
VERSION     AX029263.1  GI:10190114
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
```

```
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 3 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .12446
                     /organism="Salmonella enterica"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match             100.0%;  Score 515;  DB 6;  Length 12446;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-83;
  Matches  515;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 3420

Qy         61 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 3480

Qy        121 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 3540

Qy        181 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 3600

Qy        241 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 3660

Qy        301 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT 3720

Qy        361 ATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 ATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA 3780

Qy        421 AATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 AATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGAT 3840

Qy        481 AATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 515
              |||||||||||||||||||||||||||||||||||
Db       3841 AATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 3875


RESULT 4
AF048749
LOCUS       AF048749               24454 bp    DNA     linear   BCT 23-JUN-1999
DEFINITION  Bacteroides fragilis capsular polysaccharide C biosynthesis operon,
            complete sequence.
ACCESSION   AF048749
VERSION     AF048749.1  GI:5199106
KEYWORDS    .
SOURCE      Bacteroides fragilis
  ORGANISM  Bacteroides fragilis
            Bacteria; Bacteroidetes; Bacteroidetes (class); Bacteroidales;
            Bacteroidaceae; Bacteroides.
REFERENCE   1  (sites)
  AUTHORS   Comstock,L.E., Coyne,M.J., Tzianabos,A.O., Pantosti,A.,
            Onderdonk,A.B. and Kasper,D.L.
  TITLE     Analysis of a capsular polysaccharide biosynthesis locus of
            Bacteroides fragilis
  JOURNAL   Infect. Immun. 67 (7), 3525-3532 (1999)
   PUBMED   10377135
REFERENCE   2  (bases 1 to 24454)
  AUTHORS   Comstock,L.E. and Coyne,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1998) Channing Laboratory, Brigham & Women's
            Hospital and Harvard Medical School, 181 Longwood Avenue, Boston,
            MA 02115-5804, USA
FEATURES             Location/Qualifiers
     source          1. .24454
                     /organism="Bacteroides fragilis"
                     /mol_type="genomic DNA"
                     /strain="NCTC 9343"
                     /db_xref="taxon:817"
     CDS             complement(<1. .966)
                     /note="orf1"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAD40705.1"
                     /db_xref="GI:5199107"
                     /translation="MNNVQQVKTYSQRKISDFLFILWAGGAALLSYSLVYALRKPYTA
                     AGFDGLEAFGMDYKVVVTIAQILGYVLSKFIGIKLISELKRENRMKFILISIILAEAS
                     LILFGLLPAPYNTGAMFLNGLSLGCMWGIIFSFIEGRRMTDILASLLGVSMVISSGTA
                     KSAGLYVMDTLNVSEFWMPALIGGVALPLLALLGYALNRLPQPTAEDIAMKSKRETLN
                     GKQRWELFKNFMPFLTLLFIANVVLTILRDIKEDFLVKIIDVSQYSSWMFAQVDSVVT
                     LIILIIFGLMVFVRSNLKALSILLGLIIASMVVMAVVSFGYEQLQL"
     CDS             1272. .1907
                     /note="orf2"
                     /codon_start=1
                     /transl_table=11
                     /product="putative methyl transferase"
                     /protein_id="AAD40706.1"
                     /db_xref="GI:5199108"
                     /translation="MNKRGFVSRILQNFRKPEGFFGRMILWGMNTGHASLAQWGMSCL
                     QWQPEWSVLDIGCGGGANLLQILQRCPQGKAYGIDISPESVTFARKKNKKYLGTRCFI
                     EQGGVHRLPYPDYAFDAVTAFETVYFWGNLQHAFTEVARVLKPGGSFLICCEISDPAN
                     KAWTGLVEGMEIHSCDELKAILSKSGFTDTAIFRTKKEELCLVSHRQTVRL"
     gene            2798. .3139
                     /gene="upcY"
     CDS             2798. .3139
                     /gene="upcY"
                     /note="formerly orf3"
                     /codon_start=1
                     /transl_table=11
                     /product="UpcY"
                     /protein_id="AAD40707.1"
                     /db_xref="GI:5199109"
                     /translation="MVFVHISRQEMAAVRSTLNVYDFLRDRSTGAPTCIPDAQMADFR
                     YMLDYSQDQVILTGESIPKGTRVVVAKGDLQGLRGELVRYNNKYHILVRIDMFGSAMV
                     TIPASYVRKEK"
     gene            3322. .3714
                     /gene="upcZ"
     CDS             3322. .3714
                     /gene="upcZ"
                     /note="formerly orf4"
                     /codon_start=1
                     /transl_table=11
                     /product="UpcZ"
                     /protein_id="AAD40708.1"
                     /db_xref="GI:5199110"
                     /translation="MKNSLCNSVSSVVKKLLEYGEDGTPVYVNELTALNQELRNLCAD
                     LLLQKGESPEEEAEILVTLFKGYDTMLFNFSSENEQVIQELLDRSMTVLEKLPASVLK
                     CQLLLECFEQTGDEELIREAKKNIERVI"
     gene            <3717. .19221
                     /gene="capsular polysaccharide C biosynthesis operon"
                     /note="transcriptional start has not been mapped so first
                     gene in operon is unknown"
     gene            3717. .4604
                     /gene="rmlA"
     CDS             3717. .4604
                     /gene="rmlA"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose-1-phosphate thymidyl transferase"
                     /protein_id="AAD40709.1"
```

```
                     /db_xref="GI:5199111"
                     /translation="MKGIVLAGGSGTRLYPITKGVSKQLLPIFDKPMIYYPISVLMLA
                     GIREILIISTPDDLPGFQRLLGDGSDFGVRFEYAEQPSPDGLAQAFIIGEKFIGDDSV
                     CLVLGDNIFYGQGFTRMLNEAVRIAESESKATVFGYWVSDPERYGVAEFDENGNVFSI
                     EEKPQKPKSNYAVVGLYFYPNKVVEVAKSIRPSSRGELEITTVNQNFLSDKELRVQLL
                     GRGFAWLDTGTHDSLSEASTFIEVIEKRQGLKVACLEGIALRKGWISPEKMKALAQPM
                     LKNQYGQYLLKVIDELYVK"
gene                 4620. .5168
                     /gene="rmlC"
CDS                  4620. .5168
                     /gene="rmlC"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-6-deoxy-D-glucose-3,5 epimerase"
                     /protein_id="AAD40710.1"
                     /db_xref="GI:5199112"
                     /translation="MNIIKTAIEGLVILEPQLFKDDRGYFFESFNQREFEEKVCKTTF
                     VQDNESKSSYGVIRGLHFQKPPFAQSKLVRVIKGAVLDVAVDIRKGSPTFGKHVSVEL
                     TEDNHRQFFIPRGFAHGFSVLSEEVIFQYKCDNFYHPEAEGAIAWNDPDLGIEWRVPC
                     KSIILSKKDRVHPLLKYIILFK"
gene                 5187. .5708
                     /gene="wcfA"
CDS                  5187. .5708
                     /gene="wcfA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative acetyl transferase"
                     /protein_id="AAD40711.1"
                     /db_xref="GI:5199113"
                     /translation="MKKLVYDVRKDIIPWNSRITRILFFLNKIPYLRLFSKPLLRKQF
                     NLSNTVQFNSGFFCHAPKLKCGNYVGLSDTFILAYADVVIGNNVSFSFRNMLITSTHD
                     VNDFNKIIASSIVIGNNVWITSNVIILAGVKIGDNTIIGAGSVVTHDIPANVFAAGNP
                     CRVIKSIRFNKNE"
gene                 5701. .7236
                     /gene="wzx"
CDS                  5701. .7236
                     /gene="wzx"
                     /codon_start=1
                     /transl_table=11
                     /product="putative flippase"
                     /protein_id="AAD40712.1"
                     /db_xref="GI:5199114"
                     /translation="MNNSRTAKGLKNAQVALLFYFINLILSFLSRKAFIEHLGAEVLG
                     LNTTLTNILVYLNLAELGIGYAISYALYKPLYEGDKQAVNEIISIQGWLYKRIAIIVI
                     IMSGVILLFFPLFFAKIEIPIWYVYATFLVLLSSSIFSYFFNYKQFVLTADQKEYKLI
                     TNVQGFKVVKTILQIVVIVYFCNGYVWWLLLELLMGVITVFVLNSIVRKEYPWLQTSP
                     KIGKDVKDKYPYIIKKTKQLFFHKIANVVLNQTSPIIIYSYTNLTMVAVYGNYMLIIS
                     GISLLINSVFSSIGAGIGNLVAEGNQAKIIQVFNELLSSRIWIVSILCFGVYQMSRPF
                     IVLWVGDRFVLDDFYLLLMLLIAFISLTRLVDLFIAAYGLYQDVWAAILEAFLNLGLS
                     ILFGYYWGLSGILGGVIVSLVIIALLWKPFFLFKYGFNKNCLNFYFVYMKCVAFALIT
                     FYFSIRVIDYIGIGMCTDYSSWILISMLNIFVYTIISFPIFFFFSDGTNRFIKRVINI
                     VFN"
gene                 7240. .8103
                     /gene="wcfB"
CDS                  7240. .8103
                     /gene="wcfB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative fucosyl transferase"
                     /protein_id="AAD40713.1"
                     /db_xref="GI:5199115"
                     /translation="MLYVILRGRLGNNLFQIATAASLTQNFIFCTVNKDQERQVLLYK
                     DSFFKNIKVMKGVPDGIPYYKEPFHEFSRIPYEEGKDLIIDGYFQSEKYFKRSVVLDL
                     YRITDELRKKIWNICGNILEKGETVSIHVRRGDYLKLPHALPFCGKSYYKNAIQYIGE
                     DKIFIICSDDIDWCKKNFIGKRYYFIENTTPLLDLYIQSLCTHNIISNSSFSWWGAWL
                     NENSNKIVIAPQMWFGISVKLGVSDLLPVSWVRLPNNYTLGRYCFALYKVVEDYLLNI
                     LRLIWKRKKNM"
gene                 8079. .9239
                     /gene="wcfC"
CDS                  8079. .9239
                     /gene="wcfC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glycosyl transferase"
                     /protein_id="AAD40714.1"
                     /db_xref="GI:5199116"
                     /translation="MEKKEEYVKKKMTLIVNHPWNDVFEGKDVFLVPCYIGKIYHYDV
                     KIVFPSNYLYEAKTIRDVELVPVSYIKKLRPFSFVFGFKLMRYLFCNARKIDLFMRFH
                     VTITTAIMIIIYKMLNRNGIAYIKLDSNGVMNFEYKNNLKSLFRKLLYRRMFELVDFV
                     SYETKMGLENITQQTIGIDISSKLFYMPNGFDEDMIKHLDINVKEYSDKENVMITVGR
                     LGTNQKNTELFLKAVENIDLKDWVIYLIGDIDPQNITPLEFVSNFFSNHPEKKKSVIF
                     TGAISDKRHLWEYYNKSKVFVLTSRCESYGLVLNEAKRFRNFIVSTNVGAFEDLVESG
                     KYGCEIPQDNTDYLACILEKIILGQLDIDVYNDFSPESLSYYYQVKRMKLNN"
gene                 9257. .9913
                     /gene="wcfD"
CDS                  9257. .9913
                     /gene="wcfD"
                     /codon_start=1
                     /transl_table=11
                     /product="putative acetyl transferase"
                     /protein_id="AAD40715.1"
                     /db_xref="GI:5199117"
                     /translation="MIYYLFRPCLYVCKMFLNYFVCFTSKICTENRYNYFKRIKPFLF
                     SRIVRTWFGHVGDKISIVPPFYIRGGKYIQLGTNFRALSGLRLEAIDFYNNSSYHPSI
                     VIGDNVVFNNSCHIGATNRIIIGNDVVVASRVFITDHFHGNTTYNDLQIPVRNRLLYS
                     KGPVVIGNNVWIGEGASILPNVTIGDNSIVAAHAVVTKNIPANSIAVGCPARIIRTIN
                     "
gene                 9914. .11011
                     /gene="wzy"
CDS                  9914. .11011
                     /gene="wzy"
                     /codon_start=1
                     /transl_table=11
                     /product="putative polymerase"
                     /protein_id="AAD40716.1"
                     /db_xref="GI:5199118"
                     /translation="MNLFVLLFVFVFGIIGSYSSKFSEKNLFWLFVIVLAIIIGGRSI
                     SVPDTDQYYQLTLSSSILDIPLYSYEIGFQLLTKILKLITNGSSFFYFGIIVLINSYF
                     LFKSFSCFTDQSYPSKKSSGTLINQIKLQEFLLLYFSYWGLYYTGIVLRAGIAMSILI
```

```
 Query Match             11.5%;  Score 59.4;  DB 1;  Length 24454;
 Best Local Similarity   52.2%;  Pred. No. 0.14;
 Matches  157;  Conservative    0;  Mismatches  141;  Indels    3;  Gaps    1;

Qy        215 TTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGTGATC 274
              | |||||  |  ||  ||||  | ||       | | | ||   | |    || || |||
Db       9516 TGGAAGCAATTGATTTTTATAATAATAGTTCTTATCATCCAAGTATAGTAATTGGGGATA 9575

Qy        275 GTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAACG 334
              | |   ||| | | ||| |  |||||||              |  | ||| | |  ||||
Db       9576 ATGTTGTTTTTAATAATAGTTGTCATATTGGAGCTACTAATAGAATTATTATTGGTAACG 9635

Qy        335 ATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAAGTAT 394
              || || |  | |  ||   |||||  |||   ||||||   ||||| |   ||| |   |
Db       9636 ATGTAGTGGTGGCAAGTCGAGTTTTCATAACAGATCATTTTCATGGTA---ATACAACGT 9692

Qy        395 GCAGTCCGAAAATAGAACCGCCAGCAAATAAGCCATTAGGTGATATTGCTCCTATTAAAA 454
              || |     |  |   ||  | ||| |  | | ||  |  ||  | |||| ||   |
Db       9693 ACAATGATTTACAAATTCCAGTAAGAAACAGACTACTATATTCTAAAGGTCCTGTTGTTA 9752

Qy        455 TAGGTAATTGCTGCTGGATTGGAGATAATGCAGTAATTCTGGCTGGTAGTGAAATTTGTG 514
              ||||||||      ||||||||||     ||    |||| || |   |  |  |||| | |
Db       9753 TAGGTAATAATGTATGGATTGGAGAAGGAGCTTCAATTTTGCCAAATGTTACAATTGGAG 9812

Qy        515 A 515
              |
Db       9813 A 9813


RESULT 5
CR626927_12
WPCOMMENT
  Sequence split into 52 fragments  LOCUS CR626927 Accession CR626927
     Fragment Name      Begin     End
     CR626927_00            1    110000
     CR626927_01       100001    210000
```

CR626927_02	200001	310000
CR626927_03	300001	410000
CR626927_04	400001	510000
CR626927_05	500001	610000
CR626927_06	600001	710000
CR626927_07	700001	810000
CR626927_08	800001	910000
CR626927_09	900001	1010000
CR626927_10	1000001	1110000
CR626927_11	1100001	1210000
CR626927_12	1200001	1310000
CR626927_13	1300001	1410000
CR626927_14	1400001	1510000
CR626927_15	1500001	1610000
CR626927_16	1600001	1710000
CR626927_17	1700001	1810000
CR626927_18	1800001	1910000
CR626927_19	1900001	2010000
CR626927_20	2000001	2110000
CR626927_21	2100001	2210000
CR626927_22	2200001	2310000
CR626927_23	2300001	2410000
CR626927_24	2400001	2510000
CR626927_25	2500001	2610000
CR626927_26	2600001	2710000
CR626927_27	2700001	2810000
CR626927_28	2800001	2910000
CR626927_29	2900001	3010000
CR626927_30	3000001	3110000
CR626927_31	3100001	3210000
CR626927_32	3200001	3310000
CR626927_33	3300001	3410000
CR626927_34	3400001	3510000
CR626927_35	3500001	3610000
CR626927_36	3600001	3710000
CR626927_37	3700001	3810000
CR626927_38	3800001	3910000
CR626927_39	3900001	4010000
CR626927_40	4000001	4110000
CR626927_41	4100001	4210000
CR626927_42	4200001	4310000
CR626927_43	4300001	4410000
CR626927_44	4400001	4510000
CR626927_45	4500001	4610000
CR626927_46	4600001	4710000
CR626927_47	4700001	4810000
CR626927_48	4800001	4910000
CR626927_49	4900001	5010000
CR626927_50	5000001	5110000
CR626927_51	5100001	5205140

Continuation (13 of 52) of CR626927 from base 1200001 (CR626927 Bacteroides fragilis NCT

Query Match 11.5%; Score 59.4; DB 1; Length 110000;
Best Local Similarity 52.2%; Pred. No. 0.077;
Matches 157; Conservative 0; Mismatches 141; Indels 3; Gaps 1;

```
Qy        215 TTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGTGATC 274
              | |||||  |  ||  ||||  | ||      | | | ||   | |    || || |||
Db      67810 TGGAAGCAATTGATTTTTATAATAATAGTTCTTATCATCCAAGTATAGTAATTGGGGATA 67869

Qy        275 GTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAACG 334
              | |   ||| | | ||| |  |||||||           |  | ||| | |  ||||
Db      67870 ATGTTGTTTTTAATAATAGTTGTCATATTGGAGCTACTAATAGAATTATTATTGGTAACG 67929

Qy        335 ATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAAGTAT 394
              || || |  | |  ||   |||||  |||   ||||||  |||||  |   ||| |    |
Db      67930 ATGTAGTGGTGGCAAGTCGAGTTTTCATAACAGATCATTTTCATGGTA---ATACAACGT 67986

Qy        395 GCAGTCCGAAAATAGAACCGCCAGCAAATAAGCCATTAGGTGATATTGCTCCTATTAAAA 454
              || |      |  |   ||   |  ||| |  | | ||  |  ||  | |||| ||   |
Db      67987 ACAATGATTTACAAATTCCAGTAAGAAACAGACTACTATATTCTAAAGGTCCTGTTGTTA 68046

Qy        455 TAGGTAATTGCTGCTGGATTGGAGATAATGCAGTAATTCTGGCTGGTAGTGAAATTTGTG 514
              ||||||||      ||||||||||    ||   |||| || |   |  |  |||| | |
Db      68047 TAGGTAATAATGTATGGATTGGAGAAGGAGCTTCAATTTTGCCAAATGTTACAATTGGAG 68106

Qy        515 A 515
              |
Db      68107 A 68107
```

RESULT 6
CP000076_58
WPCOMMENT
Sequence split into 71 fragments LOCUS CP000076 Accession CP000076

Fragment Name	Begin	End
CP000076_00	1	110000
CP000076_01	100001	210000
CP000076_02	200001	310000
CP000076_03	300001	410000
CP000076_04	400001	510000
CP000076_05	500001	610000
CP000076_06	600001	710000
CP000076_07	700001	810000
CP000076_08	800001	910000
CP000076_09	900001	1010000
CP000076_10	1000001	1110000
CP000076_11	1100001	1210000
CP000076_12	1200001	1310000
CP000076_13	1300001	1410000
CP000076_14	1400001	1510000
CP000076_15	1500001	1610000
CP000076_16	1600001	1710000
CP000076_17	1700001	1810000
CP000076_18	1800001	1910000
CP000076_19	1900001	2010000
CP000076_20	2000001	2110000
CP000076_21	2100001	2210000
CP000076_22	2200001	2310000
CP000076_23	2300001	2410000
CP000076_24	2400001	2510000
CP000076_25	2500001	2610000
CP000076_26	2600001	2710000
CP000076_27	2700001	2810000
CP000076_28	2800001	2910000
CP000076_29	2900001	3010000
CP000076_30	3000001	3110000
CP000076_31	3100001	3210000
CP000076_32	3200001	3310000
CP000076_33	3300001	3410000
CP000076_34	3400001	3510000
CP000076_35	3500001	3610000
CP000076_36	3600001	3710000
CP000076_37	3700001	3810000
CP000076_38	3800001	3910000
CP000076_39	3900001	4010000
CP000076_40	4000001	4110000
CP000076_41	4100001	4210000
CP000076_42	4200001	4310000
CP000076_43	4300001	4410000
CP000076_44	4400001	4510000
CP000076_45	4500001	4610000
CP000076_46	4600001	4710000
CP000076_47	4700001	4810000
CP000076_48	4800001	4910000
CP000076_49	4900001	5010000
CP000076_50	5000001	5110000
CP000076_51	5100001	5210000
CP000076_52	5200001	5310000
CP000076_53	5300001	5410000
CP000076_54	5400001	5510000
CP000076_55	5500001	5610000
CP000076_56	5600001	5710000
CP000076_57	5700001	5810000
CP000076_58	5800001	5910000
CP000076_59	5900001	6010000

```
     CP000076_60       6000001   6110000
     CP000076_61       6100001   6210000
     CP000076_62       6200001   6310000
     CP000076_63       6300001   6410000
     CP000076_64       6400001   6510000
     CP000076_65       6500001   6610000
     CP000076_66       6600001   6710000
     CP000076_67       6700001   6810000
     CP000076_68       6800001   6910000
     CP000076_69       6900001   7010000
     CP000076_70       7000001   7074893
Continuation (59 of 71) of CP000076 from base 5800001 (CP000076 Pseudomonas fluorescens

  Query Match             11.2%;  Score 57.8;  DB 1;  Length 110000;
  Best Local Similarity   55.8%;  Pred. No. 0.15;
  Matches  110;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        197 TTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTT 256
              | |||||| |||  ||| | |||||  | |    | ||| ||||  | |  ||||| ||
Db      63522 TGGGTGATCACTTTTGGTTGGAAGCCATCTCAGAGTATCGTGACCAGATTTTTGAGCCAC 63581

Qy        257 ATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCAT 316
              |  |  | || || |||  | | |    ||| ||   ||||||  || | |||||
Db      63582 ACATTGAGATCGGAGATAATTTTTCCGCAAGCGACTTTGTTCACCTTGCCTGCGTTAACA 63641

Qy        317 GTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCC 376
                |    ||| |    |  || | | || ||||| ||  | |||   || |||   |
Db      63642 AAATCTCCATAGGACGAAACGTACTTTTAGGAAGCAAGGTGCACATAACAGACCATTCTC 63701

Qy        377 ATGGCAGTTATAAAGTA 393
              ||||   |||||  | |
Db      63702 ATGGAGTTTATAGCGGA 63718


RESULT 7
BX323062/c
LOCUS       BX323062              199133 bp    DNA     linear   VRT 14-JUL-2004
DEFINITION  Zebrafish DNA sequence from clone DKEYP-38H2 in linkage group 5,
            complete sequence.
ACCESSION   BX323062
VERSION     BX323062.10  GI:50300002
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 199133)
  AUTHORS   Hunter,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 14, 2004 this sequence version replaced gi:50234193.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEYP-38H2
            is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5.
FEATURES             Location/Qualifiers
     source          1. .199133
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-38H2"
                     /clone_lib="DanioKeypilot"
ORIGIN

  Query Match             10.6%;  Score 54.6;  DB 5;  Length 199133;
  Best Local Similarity   47.5%;  Pred. No. 0.45;
  Matches  162;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy          5 TTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGC 64
              ||||    | || ||  ||  ||  |  || |  ||||   ||||| | |   |  | ||
Db     129192 TTGTGGTCATTTAAATCTTTGAACCAGGTCATTACATACTGTTATTCACTGTAAAAATGC 129133

Qy         65 GGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGA 124
              |   || | |  |    ||| |||| | |  |    | |||||  |   | |||   |
Db     129132 TGGGTTCCACAAAATTTATTCCTGTTGTTCCAACACAAATTGATTGAAGTGTTGTTACAA 129073

Qy        125 AGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGG 184
              |     ||  ||||||  | ||||  |    |  ||||||   ||  |    || ||
Db     129072 ATTTAATTGGATTGAACATAAAACAATTAAGTATAAAAAAACTCAATTACATCAGATCAT 129013

Qy        185 GTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAA 244
              |   ||  |   |  | | ||          |    |  ||  || |||
Db     129012 TTTAAATAAGTAATTTTAAAAAGCAGCAATTATTTTTTAAATGTTGTTTTTGTTATGGGG 128953

Qy        245 AATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTT 304
              |||  ||  | |  |||  || |   |||   |||||  || |  | | | |   |
Db     128952 TATTAAAATGTAACCTGTGTATGGAGTATCTGCTATGTGAAATTATTGAAGATGGCAAAA 128893

Qy        305 CTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATT 345
              | || | |  | | | |||||   |||    |||  | ||
Db     128892 ATGGCCTGTACTTTGTTATTTTGTTAAAATGAATAACATTT 128852


RESULT 8
AX344563
LOCUS       AX344563              349980 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 14 from Patent WO0200932.
ACCESSION   AX344563
VERSION     AX344563.1  GI:18492449
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of known genetic parameters within the mhc
  JOURNAL   Patent: WO 0200932-A 14 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="synthetic construct"
```

```
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)
                     Original length of seq 1: 3.673778 <223>
                     split as follows.
                     seq 01 0.000.001 TO 0.349.980
                     seq 02 0.300.001 649.980
                     seq 03 600.001 949.980
                     seq 04 900.001 1.249.980
                     seq 05 1.200.001 1.549.980
                     seq 06 1.500.001 1.849.980
                     seq 07 1.800.001 2.149.980
                     seq 08 2.100.001 2.449.980
                     seq 09 2.400.001 2.749.980
                     seq 10 2.700.001 3.049.980
                     seq 11 3.000.001 3.349.980
                     seq 12 3.300.001 3.649.980
                     seq 13 3.600.001 3.673.778 <223>
                     Original length of seq 2: 3.673778 <223>
                     split as follows.
                     seq 14 0.000.001 TO 0.349.980
                     seq 15 0.300.001 649.980
                     seq 16 600.001 949.980
                     seq 17 900.001 1.249.980
                     seq 18 1.200.001 1.549.980
                     seq 19 1.500.001 1.849.980
                     seq 20 1.800.001 2.149.980
                     seq 21 2.100.001 2.449.980
                     seq 22 2.400.001 2.749.980
                     seq 23 2.700.001 3.049.980
                     seq 24 3.000.001 3.349.980
                     seq 25 3.300.001 3.649.980
                     seq 26 3.600.001 3.673.778"
ORIGIN

  Query Match             10.3%;  Score 53.2;  DB 6;  Length 349980;
  Best Local Similarity   49.5%;  Pred. No. 0.65;
  Matches  191;  Conservative    0;  Mismatches  193;  Indels    2;  Gaps    2;

Qy         12 TAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATC 71
              ||   |||  | |   | |  | |||| | |  ||||||   || |  |  |  | | |
Db      45704 TATAGTAAATTAAGGTTAATTTATTATAAAAGAATTATTTTTTATAAATTTGATGTAGTT 45763

Qy         72 AAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTT 131
               ||    |   |  || |   | |||  | |    ||  |||  |  | ||     || |
Db      45764 TAATTGTATTATGTTTTTAAAGTTTATAGTAGTGTATAGTAAGGTTTTAGGTTTTTATAT 45823

Qy        132 TAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTAT 191
              | | ||     | |   ||| | |    | | |||    | ||  ||   |  |  | |
Db      45824 TTATTTATTATTTATTTATTAATTTATTTAGAGTAATTTTTATTTAAATATGTTTTATTT 45883

Qy        192 GCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGA 251
                ||||  ||  |   |   ||  ||    | | ||| |||    | ||     | |||
Db      45884 AATGGTTAGTATTTTGTATAGGTGTGTTATTTTTTATTATTTATATTATATTTTAATTGT 45943

Qy        252 ACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATT-TCTTGCG 310
              |  || ||| |   |||||      |||||  | ||| || | ||  |  | | || | ||
Db      45944 ATTTTTTTTATGTTTAGGTATGTTTATATATATAAATGGTTATTATTGTGTTATGTTATT 46003

Qy        311 TATCATGTTTAATTTTAG-AAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGAT 369
              ||| ||||||  | | ||| || ||| |||||   || || |   |      |||||  ||
Db      46004 TATTATGTTTTAGTATAGTAATACGTTATATAGGTTTGTTGTTTAAGAATAATAGGTTAT 46063

Qy        370 CATAGCCATGGCAGTTATAAAGTATG 395
              | |||  |    || | |  |||| |
Db      46064 CGTAGATAGTTTAGGTGTGTAGTAGG 46089


RESULT 9
AF275222
LOCUS       AF275222               1589 bp    DNA     linear   INV 12-MAR-2002
DEFINITION  Schlechtendalia chinensis 12S small subunit ribosomal RNA gene,
            partial sequence; tRNA-Val gene, complete sequence; and 16S large
            subunit ribosomal RNA gene, partial sequence; mitochondrial genes
            for mitochondrial products.
ACCESSION   AF275222
VERSION     AF275222.1  GI:13160544
KEYWORDS    .
SOURCE      mitochondrion Schlechtendalia chinensis (horned gall aphid)
  ORGANISM  Schlechtendalia chinensis
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
            Aphidoidea; Pemphigidae; Melaphidinae; Schlechtendalia.
REFERENCE   1  (bases 1 to 1589)
  AUTHORS   von Dohlen,C.D. and Moran,N.A.
  TITLE     Molecular data support a rapid radiation of aphids in the
            cretaceous and multiple origins of host alternation
  JOURNAL   Biol. J. Linn. Soc. Lond. 71 (4), 689-717 (2000)
REFERENCE   2  (bases 1 to 1589)
  AUTHORS   von Dohlen,C.D. and Moran,N.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2000) Department of Biology, Utah State
            University, UMC 5305, Logan, UT 84322, USA
FEATURES             Location/Qualifiers
     source          1. .1589
                     /organism="Schlechtendalia chinensis"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:38111"
     rRNA            <1. .366
                     /product="12S small subunit ribosomal RNA"
     tRNA            367. .462
                     /product="tRNA-Val"
     rRNA            463. .>1589
                     /product="16S large subunit ribosomal RNA"
ORIGIN

  Query Match             10.3%;  Score 53;  DB 2;  Length 1589;
  Best Local Similarity   47.7%;  Pred. No. 5.5;
  Matches  218;  Conservative    0;  Mismatches  235;  Indels    4;  Gaps    2;

Qy          6 TGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCG 65
              | | || |||| |   |   | | |  | || |  ||| | |||| |  |||
Db        639 TTTATAGAGTTAATTTTATTATTAATTTATTGTTTTATAAATATTTAAGAAATAATTTTA 698

Qy         66 GAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAA 125
              |   | |||   ||      |  | || |  ||||   || | |  || ||||       ||
Db        699 GTTTTGAAATTTAATTCGTTATATTATATTTAG-TTATTTAAGATTTAAATTTAATTTAA 757

Qy        126 GAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGG 185
              | ||| ||| |  || |  ||  ||| |     ||| | |       |      | |   |
Db        758 AATTTTAAAAGTTTATTTATAATTTTATTTATAAAATATTTTAGATTTAAATTATTTTAG 817

Qy        186 TAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAA 245
                || ||  |   |  ||  |  | |    |   || |||||| |  ||  | ||
Db        818 GGATAAGCTTGAAATTTTTATATAATAGTAAAATTTATTAAATAATTTTTAGGAATAGAA 877

Qy        246 ATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTC 305
              ||||  |  | | ||     | |  | ||  | |||||||  | ||     |  | |||
Db        878 ATTTTTATTTAAGTTTGTAAAAATTTATTATTTATGTTTATTTAATTTATTATAAATTAT 937

Qy        306 TTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGG 365
              ||   | |  | || ||||||| |||   |||  ||| ||      | | |   || |
Db        938 TTATTTTTAGTATTAAATTTTAAAAATTAATAATTTATTTTTAGTAATAATGATTAAATT 997

Qy        366 CGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCAAATAA 425
              |   |  |  |   | ||||||| |||| | || | ||| |  |   || |  | |
Db        998 AGTATAATGGATT---ATTTATAAATTATGAATTCGGCAAAAATTATTTTCATCTGTTTA 1054

Qy        426 GCCATTAGGTGATATTGCTCCTATTAAAATAGGTAAT 462
              | |  |  |  ||||     |  |  ||||   |||
Db       1055 TCAAAAACATAGTATTAAGATTTATTTAATATAGAAT 1091


RESULT 10
PFMAL13_18
```

```
WPCOMMENT
  Sequence split into 28 fragments  LOCUS PFMAL13 Accession AL844509
     Fragment Name      Begin     End
     PFMAL13_00             1    110000
     PFMAL13_01        100001    210000
     PFMAL13_02        200001    310000
     PFMAL13_03        300001    410000
     PFMAL13_04        400001    510000
     PFMAL13_05        500001    610000
     PFMAL13_06        600001    710000
     PFMAL13_07        700001    810000
     PFMAL13_08        800001    910000
     PFMAL13_09        900001   1010000
     PFMAL13_10       1000001   1110000
     PFMAL13_11       1100001   1210000
     PFMAL13_12       1200001   1310000
     PFMAL13_13       1300001   1410000
     PFMAL13_14       1400001   1510000
     PFMAL13_15       1500001   1610000
     PFMAL13_16       1600001   1710000
     PFMAL13_17       1700001   1810000
     PFMAL13_18       1800001   1910000
     PFMAL13_19       1900001   2010000
     PFMAL13_20       2000001   2110000
     PFMAL13_21       2100001   2210000
     PFMAL13_22       2200001   2310000
     PFMAL13_23       2300001   2410000
     PFMAL13_24       2400001   2510000
     PFMAL13_25       2500001   2610000
     PFMAL13_26       2600001   2710000
     PFMAL13_27       2700001   2732359
Continuation (19 of 28) of PFMAL13 from base 1800001 (AL844509 Plasmodium falciparum 3D7

  Query Match           10.1%;  Score 52;  DB 14;  Length 110000;
  Best Local Similarity   46.4%;  Pred. No. 1.7;
  Matches  169;  Conservative    0;  Mismatches  195;  Indels    0;  Gaps    0;

Qy         15 TTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAA 74
              | ||| |||     ||   | |    | ||| |||||| | |   | | |   |  | ||    |
Db      15835 TATAATATTTTTTTTTTTTTTATGCGTATTATTATTCATTTTAATGATCCATATATGGTA 15894

Qy         75 AGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAA 134
              |  ||  ||| || |  |  |||      |  || |  | ||  |  | |   || |  |
Db      15895 ATTAATTTTTTTTTTCTTTTTTATTTTTAAATATATATATATTATATGTATTTATATATA 15954

Qy        135 ATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCG 194
              ||  ||    | ||   |    | |||  | ||      | ||  | |||   || | |
Db      15955 ATATAAAAATATACAAGAAAATACAATTATTACATAATATTATATATGTTATATTTTATT 16014

Qy        195 GGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACC 254
              ||  |          |  | |     | | || | | |    |||    |  |   |
Db      16015 TATTTATTTATTTATATATAATTTTTTTTTTTAAATTAAATAAGATATTTATATAATAAT 16074

Qy        255 TTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATC 314
              |   || || |||  |      ||| | ||  | ||| | | | ||||| ||   ||||
Db      16075 TATATTATATATATATATAATATTATATATATTTTATATTTTACTTATTTATTTAATATC 16134

Qy        315 ATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAG 374
              | || | || | ||   ||||   ||| |  | | | | | ||||    || |||
Db      16135 TATATAAAATATATATAATATTATACATATTTGCTGTATATTATTTATAATATATTATAT 16194

Qy        375 CCAT 378
              ||
Db      16195 ATAT 16198


RESULT 11
CR792440
LOCUS       CR792440              320170 bp    DNA     linear   HTG 11-FEB-2005
DEFINITION  Danio rerio clone CH211-264J2, *** SEQUENCING IN PROGRESS ***, 37
            unordered pieces.
ACCESSION   CR792440
VERSION     CR792440.3  GI:52694505
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 320170)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 25, 2004 this sequence version replaced gi:52673607.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC264J2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 305175 bases at least Q40
            Consensus quality: 308510 bases at least Q30
            Consensus quality: 310534 bases at least Q20
            Insert size: 316570; sum-of-contigs
            Insert size: 145631; 20.5% error; agarose-fp
            Quality coverage: 4.86x in Q20 bases; sum-of-contigs Quality
            coverage: 10.94x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3761: contig of 3761 bp in length
            *     3762     3861: gap of 100 bp
            *     3862    10587: contig of 6726 bp in length
            *    10588    10687: gap of 100 bp
            *    10688    18303: contig of 7616 bp in length
            *    18304    18403: gap of 100 bp
            *    18404    22685: contig of 4282 bp in length
            *    22686    22785: gap of 100 bp
            *    22786    26981: contig of 4196 bp in length
            *    26982    27081: gap of 100 bp
            *    27082    33547: contig of 6466 bp in length
            *    33548    33647: gap of 100 bp
            *    33648    39255: contig of 5608 bp in length
            *    39256    39355: gap of 100 bp
            *    39356    63244: contig of 23889 bp in length
            *    63245    63344: gap of 100 bp
            *    63345    71860: contig of 8516 bp in length
            *    71861    71960: gap of 100 bp
            *    71961    75327: contig of 3367 bp in length
            *    75328    75427: gap of 100 bp
            *    75428    80081: contig of 4654 bp in length
            *    80082    80181: gap of 100 bp
            *    80182    82309: contig of 2128 bp in length
            *    82310    82409: gap of 100 bp
            *    82410    90910: contig of 8501 bp in length
            *    90911    91010: gap of 100 bp
            *    91011    97412: contig of 6402 bp in length
            *    97413    97512: gap of 100 bp
            *    97513   105983: contig of 8471 bp in length
            *   105984   106083: gap of 100 bp
            *   106084   121267: contig of 15184 bp in length
            *   121268   121367: gap of 100 bp
            *   121368   126148: contig of 4781 bp in length
            *   126149   126248: gap of 100 bp
```

```
          *   126249   131710: contig of 5462 bp in length
          *   131711   131810: gap of 100 bp
          *   131811   136833: contig of 5023 bp in length
          *   136834   136933: gap of 100 bp
          *   136934   157207: contig of 20274 bp in length
          *   157208   157307: gap of 100 bp
          *   157308   164180: contig of 6873 bp in length
          *   164181   164280: gap of 100 bp
          *   164281   183283: contig of 19003 bp in length
          *   183284   183383: gap of 100 bp
          *   183384   192926: contig of 9543 bp in length
          *   192927   193026: gap of 100 bp
          *   193027   210181: contig of 17155 bp in length
          *   210182   210281: gap of 100 bp
          *   210282   230216: contig of 19935 bp in length
          *   230217   230316: gap of 100 bp
          *   230317   233471: contig of 3155 bp in length
          *   233472   233571: gap of 100 bp
          *   233572   253843: contig of 20272 bp in length
          *   253844   253943: gap of 100 bp
          *   253944   256589: contig of 2646 bp in length
          *   256590   256689: gap of 100 bp
          *   256690   260286: contig of 3597 bp in length
          *   260287   260386: gap of 100 bp
          *   260387   263172: contig of 2786 bp in length
          *   263173   263272: gap of 100 bp
          *   263273   270528: contig of 7256 bp in length
          *   270529   270628: gap of 100 bp
          *   270629   274513: contig of 3885 bp in length
          *   274514   274613: gap of 100 bp
          *   274614   278690: contig of 4077 bp in length
          *   278691   278790: gap of 100 bp
          *   278791   284323: contig of 5533 bp in length
          *   284324   284423: gap of 100 bp
          *   284424   291121: contig of 6698 bp in length
          *   291122   291221: gap of 100 bp
          *   291222   302577: contig of 11356 bp in length
          *   302578   302677: gap of 100 bp
          *   302678   320170: contig of 17493 bp in length.
FEATURES             Location/Qualifiers
     source          1. .320170
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-264J2"
                     /clone_lib="CHORI-211"
     misc_feature    1. .3761
                     /note="assembly_fragment:00380
                     fragment_chain:1"
     misc_feature    3862. .10587
                     /note="assembly_fragment:00647
                     fragment_chain:1"
     misc_feature    10688. .18303
                     /note="assembly_fragment:00479
                     fragment_chain:1"
     misc_feature    18404. .22685
                     /note="assembly_fragment:00207
                     fragment_chain:1"
     misc_feature    22786. .26981
                     /note="assembly_fragment:00320
                     fragment_chain:1"
     misc_feature    27082. .33547
                     /note="assembly_fragment:00557
                     fragment_chain:1"
     misc_feature    33648. .39255
                     /note="assembly_fragment:00411
                     fragment_chain:1"
     misc_feature    39356. .63244
                     /note="assembly_fragment:02294
                     fragment_chain:1"
     misc_feature    63345. .71860
                     /note="assembly_fragment:00696
                     fragment_chain:1"
     misc_feature    71961. .75327
                     /note="assembly_fragment:00080
                     fragment_chain:1"
     misc_feature    75428. .80081
                     /note="assembly_fragment:00157
                     fragment_chain:1"
     misc_feature    80182. .82309
                     /note="assembly_fragment:00042
                     fragment_chain:1"
     misc_feature    82410. .90910
                     /note="assembly_fragment:00801
                     fragment_chain:2"
     misc_feature    91011. .97412
                     /note="assembly_fragment:00442
                     fragment_chain:2"
     misc_feature    97513. .105983
                     /note="assembly_fragment:01055
                     fragment_chain:2"
     misc_feature    106084. .121267
                     /note="assembly_fragment:01151
                     fragment_chain:2"
     misc_feature    121368. .126148
                     /note="assembly_fragment:00234
                     fragment_chain:2"
     misc_feature    126249. .131710
                     /note="assembly_fragment:00516
                     fragment_chain:2"
     misc_feature    131811. .136833
                     /note="assembly_fragment:00261
                     fragment_chain:2"
     misc_feature    136934. .157207
                     /note="assembly_fragment:01894
                     fragment_chain:2"
     misc_feature    157308. .164180
                     /note="assembly_fragment:00857
                     fragment_chain:2"
     misc_feature    164281. .183283
                     /note="assembly_fragment:01552
                     fragment_chain:3"
     misc_feature    183384. .192926
                     /note="assembly_fragment:00914
                     fragment_chain:3"
     misc_feature    193027. .210181
                     /note="assembly_fragment:01269
                     fragment_chain:3"
     misc_feature    210282. .230216
                     /note="assembly_fragment:02089
                     fragment_chain:3"
     misc_feature    230317. .233471

  Query Match            10.1%;  Score 52;  DB 14;  Length 320170;
  Best Local Similarity  36.1%;  Pred. No. 1.1;
  Matches  130;  Conservative    0;  Mismatches  230;  Indels    0;  Gaps    0;

Qy          4 TTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGG 63
              | | | | || |||     | |||  |    | ||    | ||     || ||  ||  |||||
Db     121168 TATTTATTTATTTTTGTGTGAAATATTGGTTTTTAGAAATATAAATGGAAAAAGATTAGG 121227

Qy         64 CGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGG 123
               |||| ||||  | |  || || || |  |||    |
Db     121228 GGTAAATAAAAAAAAAGATATTATTATTATTGTTATTTTTTNNNNNNNNNNNNNNNNNNN 121287

Qy        124 AAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGG 183

Db     121288 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 121347

Qy        184 GGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAA 243
                                  ||  |       |  |  | ||||||  |||    ||  |
Db     121348 NNNNNNNNNNNNNNNNNNNTATATATATATATATATATATATATATATATATATATA 121407

Qy        244 AAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATT 303
               |  |  |    ||| | || ||| |     |||| | ||  |    || |  ||||
Db     121408 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121467
```

```
Qy        304 TCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATA 363
              | |         || | ||  | || | |   ||||||  ||    ||    |  | ||||||
Db     121468 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121527


RESULT 12
CR848702/c
LOCUS       CR848702              184622 bp    DNA     linear   HTG 09-JUL-2005
DEFINITION  Danio rerio chromosome 9 clone CH211-225I19, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   CR848702
VERSION     CR848702.3  GI:70671221
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 184622)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Jul 9, 2005 this sequence version replaced gi:54287324.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC225I19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 183071 bases at least Q40
            Consensus quality: 183183 bases at least Q30
            Consensus quality: 183282 bases at least Q20
            Insert size: 184422; sum-of-contigs
            Insert size: 184530; 2.1% error; agarose-fp
            Quality coverage: 8.58x in Q20 bases; sum-of-contigs Quality
            coverage: 8.58x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    65160: contig of 65160 bp in length
            *    65161    65260: gap of 100 bp
            *    65261   161321: contig of 96061 bp in length
            *   161322   161421: gap of 100 bp
            *   161422   184622: contig of 23201 bp in length.
FEATURES             Location/Qualifiers
     source          1. .184622
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="9"
                     /clone="CH211-225I19"
                     /clone_lib="CHORI-211"
     misc_feature    1. .65160
                     /note="assembly_fragment:00286
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    65261. .161321
                     /note="assembly_fragment:01282
                     fragment_chain:1"
     misc_feature    161422. .184622
                     /note="assembly_fragment:00001
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             9.9%;  Score 51.2;  DB 14;  Length 184622;
  Best Local Similarity   48.1%;  Pred. No. 1.9;
  Matches  175;  Conservative    0;  Mismatches  188;  Indels    1;  Gaps    1;

Qy          1 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 60
              || | |   ||||  |||| || | | |    | | || || | ||||||   | |  |
Db      65043 ATATATTAATATATATTAAAATAATATTATTTTATAATTATTTTATTATTTTTTATATA 64984

Qy         61 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 120
                 ||  ||     |  |||  | |  |  ||     || | ||||||||||  ||
Db      64983 TATTTTTTATTTAATATTATATTTTATTTATTATTTTTATAATTTTATTTTAAAATATTT 64924

Qy        121 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 180
              |   || |||||   |||   | |  ||| | |  ||||| ||  |     |  |
Db      64923 TTTATTTATATTAAAAATAATTATATATATTATTATTTAAAAATTATAAATATATTTATT 64864

Qy        181 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTG-TATATAATTATGGTGA 239
              |    | |||     ||  | | |    |  || |   |  | |||| ||||  |
Db      64863 TTTATTTTTATTTATATTTATATTTATATATATTTTTATTTTTATTTATATTTATTTTTT 64804

Qy        240 TGAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCA 299
              |   ||  |  |    ||| | || |||  |      ||||| | ||  |    || | |
Db      64803 TATTAATATATATAAATATATATATATATATATATATATATATATATATATATATATATA 64744

Qy        300 TATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTA 359
              ||| | |         || | ||  | || | |   ||||||  ||    ||   |  | ||
Db      64743 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64684

Qy        360 TATA 363
              ||||
Db      64683 TATA 64680


RESULT 13
AF172324
LOCUS       AF172324               14263 bp    DNA     linear   BCT 29-OCT-1999
DEFINITION  Escherichia coli GalF (galF) gene, partial cds; O-antigen repeat
            unit transporter Wzx (wzx), WbnA (wbnA), O-antigen polymerase Wzy
            (wzy), WbnB (wbnB), WbnC (wbnC), WbnD (wbnD), WbnE (wbnE),
            UDP-Glc-4-epimerase GalE (galE), 6-phosphogluconate dehydrogenase
            Gnd (gnd), UDP-Glc-6-dehydrogenase Ugd (ugd), and WbnF (wbnF)
            genes, complete cds; and chain length determinant Wzz (wzz) gene,
            partial cds.
ACCESSION   AF172324
VERSION     AF172324.1  GI:5739460
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 14263)
  AUTHORS   Paton,A.W. and Paton,J.C.
  TITLE     Molecular characterization of the locus encoding biosynthesis of
            the lipopolysaccharide O antigen of Escherichia coli serotype O113
  JOURNAL   Infect. Immun. 67 (11), 5930-5937 (1999)
   PUBMED   10531250
REFERENCE   2  (bases 1 to 14263)
  AUTHORS   Paton,A.W. and Paton,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-1999) Molecular Microbiology Unit, Women's and
            Children's Hospital, King William Road, North Adelaide, S.A.,
            Australia
FEATURES             Location/Qualifiers
     source          1. .14263
                     /organism="Escherichia coli"
```

/mol_type="genomic DNA"
/strain="98NK2"
/serotype="O113:H21"
/db_xref="taxon:562"

gene <1. .801
/gene="galF"

CDS <1. .801
/gene="galF"
/function="putative regulator of UDP-Glc pyrophosphorylase"
/codon_start=1
/transl_table=11
/product="GalF"
/protein_id="AAD50483.1"
/db_xref="GI:5739461"
/translation="VDKPMIQYIVDEIVAAGIKEILLVTHASKNAVENHFDTSYELES
LLEQRVKRQLLAEVQSICPPGVTIMNVRQGEPLGLGHSILCARPAIGDNPFVVVLPDV
VIDDASADPLRYNLAAMIARFNETGRSQVLAKRMPGDLSEYSVIQTKEPLDREGKVSR
IVEFIEKPDQPQTLDSDIMAVGRYVLSADIWPELERTQPGAWGRIQLTDAIAELAKKQ
SVDAMLMTGDSYDCGKKMGYMQAFVKYGLRNLKEGAKFRKGIEKLLSE"

gene 1121. .2542
/gene="wzx"

CDS 1121. .2542
/gene="wzx"
/codon_start=1
/transl_table=11
/product="O-antigen repeat unit transporter Wzx"
/protein_id="AAD50484.1"
/db_xref="GI:5739462"
/translation="MEKNSDKLRTGVRWSAIERIITQIAQLGILLILSRILGPKAYGL
IGMLTIFIAISQTLVDSGFSSALIRRNKNSQKQYSTIFWFNLIVSLLIYAVIFFMSPH
ISAFYREKELSDILRVLSLLVIINALSIVHRARLTIMMDFKTQTKISLIAVLGSGVVS
VILAKNGWGVWALVVNNLVFATISTFLYFWFCKWLPDFIFQLKFFKVTAAFSVKLMIA
SLLNTFFDNIYQLVIGKVFYPTQVGYFTQAKNLTLLPANTYSAIIQRVTYRYFSEIKN
DRVRLRQKYEEVMRISAFIFFPLTISFSIFSNNIIVFLIGKAWEEAAKYISILSWCYV
LYPIHAISLNILNVFGRSDLFLKVEIIKKIIITVCLITTYKYGISAICMGMVFYSLLA
TYINTIHTNRIIKTNFIALIRIISPIAFISMLSATISFVICKFFDNNIIYIFGIMISL
IIYYLIYYFVDRKIVLTIWKVNK"

gene 2544. .3464
/gene="wbnA"

CDS 2544. .3464
/gene="wbnA"
/function="putative glycosyl transferase"
/codon_start=1
/transl_table=11
/product="WbnA"
/protein_id="AAD50485.1"
/db_xref="GI:5739463"
/translation="MNNETLDVSVVIPVYNCGKTIKETINSILIQTGISFEIIVIDDG
SNDETARIVNEICDERIKYFYQNNSGISVALNNGIKRAAAKYIARIDGDDIALPTRLE
TQFYMLESNPDVAMVASAVEFISDDGIVIGRSFPYTSKYFIRDVLSHHNIYAHPSVMF
RKKTILSVGGYNSLLSGYCEDYYLWMQIIKLGKIINVSVPLTQYRISSGQITDWQPSK
NYYKVMNRIINSKYIDVNLVNKLAEIKQLEKEKCKNKAVNNIREEIIKKNIFNLAFNI
LISFRVSRKFASNTICICKNITTLMKVFLT"

gene 3480. .4667
/gene="wzy"

CDS 3480. .4667
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O-antigen polymerase Wzy"
/protein_id="AAD50486.1"
/db_xref="GI:5739464"
/translation="MNTTSVICFIVFLLNPSFGCFLFLLTLFNHEKLSIFFISLFAGI
LAYNLVPYDTMDLASHYMKFSELHNMSVSDIWSDGVNIFLYLCMKCLNAAGINKEWLP
FIATFIGYYILLNALRNNIPAYWCRKEKMICFIIWLLSISFLSLSNGIRSGLSSAFFI
YAVICQRNYRHLSFIIYSVLACMMHSFAIPMVILLFSMPLILKRKMLSQRFFKIAIIM
SVSLSFLVDINSLIVFIFNNLSFLPYIDQIYNIYIIGDRWGGGASFDTNTFIARLITQ
APFFYCILVLLISKRKDDFFILGAILVSVSLFLSQFWVVSERYQYVSVLLVTLHLIFN
NEYFLLRRKIIILLYFVIQTAVLIWGLYRFRYMILPSLEFIIYPFLFCFYNTVDMNKL
IKV"

gene 4670. .5722
/gene="wbnB"

CDS 4670. .5722
/gene="wbnB"
/function="putative glycosyl transferase"
/codon_start=1
/transl_table=11
/product="WbnB"
/protein_id="AAD50487.1"
/db_xref="GI:5739465"
/translation="MRKKILHLVGTLDVGGIEKWVVNMINYELENKNELDFYLYSLDP
HKKQIVKQINLDENSITYTNKKNIFSRMYGFYCLMKKISPDVIHYHPGYSSGIYLFIS
KIFFKKINIVHSHSDRRSIDRRSSIFKKIYIFIMKFLINRLSDYKIAVSERAGKSLFY
GSFITHYCGVPDIMLSLPDIKKVSSSEFKIYHIGRNSDAKNYPFIFSIAHSLREYENV
HIYCMGAGLELLQKKSQEENLKNMHFLGFIENPLSHIYIHANVFIMPSLWEGLPLSVV
EAQKCNVPCLVSDTITKECDIGMAEFLPLDVSIWQERILNIYKEGYLMKEVNKSKFSL
EQDILFFKKIYRDARK"

gene 5719. .6258
/gene="wbnC"

CDS 5719. .6258
/gene="wbnC"
/function="putative O-acetyl transferase"
/codon_start=1
/transl_table=11
/product="WbnC"
/protein_id="AAD50488.1"
/db_xref="GI:5739466"
/translation="MIIQIIRKIFQIFNSCFYCLFYRKLSINPIKSHFSGDVYIKNNG
AIFIGDYFRSRHGVTLNVNGGSINIGKNVFFNRYVSLNCMSHIFIGNNVLIGESSKLY
DHDHDFRSGPEEKNNKFVHQPIKVGNNVWIGSNVVILKGVTIGDNSVIGAGCIVTRDV
PANSVLIQKRHSTYIKIIK"

gene 6255. .7388
/gene="wbnD"

CDS 6255. .7388
/gene="wbnD"
/function="putative glycosyl transferase"
/codon_start=1
/transl_table=11
/product="WbnD"
/protein_id="AAD50489.1"
/db_xref="GI:5739467"
/translation="MSVKNILFIDSYYNSLYGAQKSMINLALQLQEKNINVTIASMHS
GFLLEEARNNSINTYSFNFNPNILKSEHQYSGLYQKITFIFLLFFSWIRAFILHFHKL
KTSDVICVNDIRTFLYLFPLLMLTRKKNIWYIRIREEKRKLTFLFSFITKIIIFISSD
VKKNTKIRTNVLVYVLMTGFPRFNLCYKKRKDNNEIRFVTVGSLNQRKNQVETIDLFK
KIKHSTRSKCTLDIIGSYEPKDFCYYKELNSIISNDPELNSSVFIHGYCENILNLIED
YDVFLFSSLREGLPRSIIEALHAGLFVITRPVDGVNDILLYDYLGTTYTNLKNIPFEF
CDIADKWLTPEFRLKRNQYINEKFNNEVFVDNFLSVLDNIEIK"

gene 7388. .8566
/gene="wbnE"

CDS 7388. .8566
/gene="wbnE"
/function="putative glycosyl transferase"
/codon_start=1
/transl_table=11
/product="WbnE"
/protein_id="AAD50490.1"
/db_xref="GI:5739468"
/translation="MKKIAHIQLLPMLSGVQKVTLQELMILNDNEYTKYLICKEEGEL
TEECKRLGIKTHVVKDLTREINAVKDIIALYKIYKFLKANDIDIVHTHFAKTGFLGRV
AAKLAGIPLIVHTVHGFPFDSAKNKYIAFFYKVLEFISARFADIIICLHDGDKETCKK
LLHVPESKVLVLPNGIAFTEFFRLSECDKKKARTILGIPESSLVFTMVGRLWEQKNPL
LLINAAKEVINEYPSDDIIFLLIGDGFLRKEIERIAERFIYHNKIVLLGWRKDIPDIL
SCSDVFVLPSRWEGMPLAILEAQATGLPCIVSDIPGNNNLVKDGVNGYLFESGSVTQL
KDAILRMRVIEKRHEMGINAYDYVSQKHCIHKRIDNIDNLYKKMLNKKMLNKKMKLKH
"

gene 8601. .9614
/gene="galE"

CDS 8601. .9614
/gene="galE"
/codon_start=1
/transl_table=11
/product="UDP-Glc-4-epimerase GalE"
/protein_id="AAD50491.1"
/db_xref="GI:5739469"

```
                     /translation="MKVLLTGGLGYIGCHIAVELLQQGHDVVIFDNLSNSHLTVLERI
                     KAITQKKVVFIRGDLRSLSDIRKTFESSQIEAVIHLAALKSVSESVRHPLEYYINNVV
                     GSLNLIQAMSHANIKRLIFSSSATVYGQPEHLPITEQASVGKVTNPYGQTKLIIEKIL
                     ADICLADPEWQITSLRYFNPIGAHPSGMIGEDPRGMPNNILPYITQVAIRRRPYFTIL
                     GDDYPTKDGTGVRDYIHVQDLALGHIAALMNAKNKSYQVYNLGTGKGYSVLDIINSFK
                     EINNIDIPFIIEKRRPGDIAECWSSPELAKRELGWQAKFELKDMLRDSWNWQQKNPFG
                     YGN"
     gene            9674. .11080
                     /gene="gnd"
     CDS             9674. .11080
                     /gene="gnd"
                     /codon_start=1
                     /transl_table=11
                     /product="6-phosphogluconate dehydrogenase Gnd"
                     /protein_id="AAD50492.1"
                     /db_xref="GI:5739470"
                     /translation="MSKQQIGVVGMAVMGRNLALNIESRGYTVSIFNRSREKTEEVIA
                     ENPGKKLVPYYTVKEFVESLETPRRILLMVKAGAGTDAAIDSLKPYLDKGDIIIDGGN
                     TFFQDTIRRNRELSAEGFNFIGTGVSGGEESALKGPSIMPGGQKEAYELVAPILTKIA
                     AVAEDGEPCVTYIGADGAGHYVKMVHNGIEYGDMQLIAEAYSLLKGGLNLTNEELAQT
                     FTEWNNGELSSYLIDITKDIFTKKDEDGNYLVDVILDEAANKGTGKWTSQSALDLGEP
                     LSLITESVFARYISSLKDQRVAASKVLSGPQGQPAGDKAEFIEKVRRALYLGKIVSYA
                     QGFSQLRAASEEYNWDLNYGEIAKIFRAGCIIRAQFLQKITDAYAENPQIANLLLAPY
                     FKQIADDYQQALRDVVAYAVQNGIPVPTFAAAVAYYDSYRSAVLPANLIQAQRDYFGA

  Query Match             9.9%;  Score 50.8;  DB 1;  Length 14263;
  Best Local Similarity   48.3%;  Pred. No. 5.9;
  Matches  142;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy        222 TGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATG 281
              ||   || ||| | |   |     |   || | ||  |       |  |||| | |||||
Db       5862 TGGTGATTATTTTAGGAGTCGCCATGGTGTAACTCTTAACGTTAATGGTGGTAGTATAAA 5921

Qy        282 TTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATT 341
              | || |  | |||||   | |   | |       |  || | ||         ||||  ||
Db       5922 TATAGGAAAAAATGTATTTTTCAATAGATATGTTTCATTGAATTGTATGTCTCATATTTT 5981

Qy        342 AATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCC 401
               |||||||  || ||||  |||||  ||   |     |       | ||  | |       |
Db       5982 CATTGGTAATAACGTTTTAATAGGAGAAAGTTCTAAGCTATATGATCATGACCATGATTT 6041

Qy        402 GAAAATAGAACCGCCAGCAAATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAA 461
              | ||  |  |||  || ||| ||       |  || |  |   || || ||| | || ||
Db       6042 TAGAAGTGGGCCGGAAGAAAAAAATAATAAATTTGTTCATCAACCAATAAAAGTGGGCAA 6101

Qy        462 TTGCTGCTGGATTGGAGATAATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 515
                  ||||| ||   |||||  || |||||    |||  |   ||| | ||
Db       6102 CAATGTTTGGATAGGTAGTAATGTTGTGATTCTTAAAGGTGTTACTATTGGGGA 6155


RESULT 14
CR769779/c
LOCUS       CR769779              219074 bp    DNA     linear   HTG 18-SEP-2004
DEFINITION  Danio rerio clone DKEY-18C12, *** SEQUENCING IN PROGRESS ***, 5
            unordered pieces.
ACCESSION   CR769779
VERSION     CR769779.3  GI:52345314
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 219074)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 18, 2004 this sequence version replaced gi:52214022.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK18C12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 217273 bases at least Q40
            Consensus quality: 217586 bases at least Q30
            Consensus quality: 217870 bases at least Q20
            Insert size: 218674; sum-of-contigs
            Insert size: 217168; 4.7% error; agarose-fp
            Quality coverage: 8.00x in Q20 bases; sum-of-contigs Quality
            coverage: 8.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    31758: contig of 31758 bp in length
            *    31759    31858: gap of 100 bp
            *    31859    46986: contig of 15128 bp in length
            *    46987    47086: gap of 100 bp
            *    47087    62942: contig of 15856 bp in length
            *    62943    63042: gap of 100 bp
            *    63043   117588: contig of 54546 bp in length
            *   117589   117688: gap of 100 bp
            *   117689   219074: contig of 101386 bp in length.
FEATURES             Location/Qualifiers
     source          1. .219074
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-18C12"
                     /clone_lib="DanioKey"
     misc_feature    1. .31758
                     /note="assembly_fragment:00384
                     fragment_chain:1"
     misc_feature    31859. .46986
                     /note="assembly_fragment:00008
                     fragment_chain:1"
     misc_feature    47087. .62942
                     /note="assembly_fragment:00176.0"
     misc_feature    63043. .117588
                     /note="assembly_fragment:00812"
     misc_feature    117689. .219074
                     /note="assembly_fragment:01508"
ORIGIN

  Query Match             9.9%;  Score 50.8;  DB 14;  Length 219074;
  Best Local Similarity   48.3%;  Pred. No. 2.1;
  Matches  171;  Conservative    0;  Mismatches  182;  Indels    1;  Gaps    1;

Qy         10 TATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAA 69
              || | |||   ||||| | ||||| || |  |||  || || | |||  |        ||
Db     218948 TAAAATTTTTTATTAATAATAAATATTTTTTTATTTTTTTTAATTAATATATTTAATTAA 218889

Qy         70 TCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAAT 129
              |   ||  | | | | || ||    ||     ||||  |||| | |  ||   |  |||
Db     218888 TTTTAAAAATAATATTTTATTTATATTTATTTATATTTATTTATATTTATTTTAATTAAT 218829

Qy        130 TTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGT 189
              | || |||  || | ||   |||   |   || |   | |  | ||   ||   ||
Db     218828 TATATATTTTATTTAAATATTTATATTATTAATATATATATATATTTTT-ATAATTTATA 218770

Qy        190 ATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTT 249
              |     ||     ||| |       | |   | ||||||  |||    ||    |  |
Db     218769 TTTATATTTATATTTAAATATATATATTATTTTATATATATATATATATATATTTATATA 218710
```

```
Qy        250 GAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGC 309
              |    ||| | || |||  |      ||||| | ||  |    || |  |||| | |
Db     218709 TATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218650

Qy        310 GTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATA 363
              || | ||  | || | |   ||||||  ||   ||   |  ||||||||
Db     218649 TATATATATATATATATATATATATATATATATATATATATATATATATTTATATA 218596
```

RESULT 15
CR848045

```
LOCUS       CR848045              228279 bp    DNA     linear   HTG 25-JUL-2005
DEFINITION  Danio rerio chromosome 16 clone DKEY-8K3, WORKING DRAFT SEQUENCE, 9
            unordered pieces.
ACCESSION   CR848045
VERSION     CR848045.3  GI:54304228
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 228279)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Oct 19, 2004 this sequence version replaced gi:54260956.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK8K3
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 225626 bases at least Q40
            Consensus quality: 226279 bases at least Q30
            Consensus quality: 226758 bases at least Q20
            Insert size: 227479; sum-of-contigs
            Insert size: 228967; 2.4% error; agarose-fp
            Quality coverage: 7.58x in Q20 bases; sum-of-contigs Quality
            coverage: 8.23x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    43453: contig of 43453 bp in length
            *   43454    43553: gap of 100 bp
            *   43554   130339: contig of 86786 bp in length
            *  130340   130445: gap of 106 bp
            *  130446   157655: contig of 27210 bp in length
            *  157656   157755: gap of 100 bp
            *  157756   173246: contig of 15491 bp in length
            *  173247   173346: gap of 100 bp
            *  173347   180935: contig of 7589 bp in length
            *  180936   181035: gap of 100 bp
            *  181036   194156: contig of 13121 bp in length
            *  194157   194256: gap of 100 bp
            *  194257   209151: contig of 14895 bp in length
            *  209152   209251: gap of 100 bp
            *  209252   211362: contig of 2111 bp in length
            *  211363   211462: gap of 100 bp
            *  211463   228279: contig of 16817 bp in length.
FEATURES             Location/Qualifiers
     source          1. .228279
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="16"
                     /clone="DKEY-8K3"
                     /clone_lib="DanioKey"
     misc_feature    1. .43453
                     /note="assembly_fragment:01348
                     fragment_chain:1"
     misc_feature    43554. .130339
                     /note="assembly_fragment:02021
                     fragment_chain:1"
     misc_feature    130440. .157655
                     /note="assembly_fragment:00953
                     fragment_chain:1"
     misc_feature    157756. .173246
                     /note="assembly_fragment:00357
                     fragment_chain:1"
     misc_feature    173347. .180935
                     /note="assembly_fragment:00148
                     fragment_chain:1"
     misc_feature    181036. .194156
                     /note="assembly_fragment:00212
                     fragment_chain:2"
     misc_feature    194257. .209151
                     /note="assembly_fragment:00544
                     fragment_chain:2"
     misc_feature    209252. .211362
                     /note="assembly_fragment:00142"
     misc_feature    211463. .228279
                     /note="assembly_fragment:00738.0"
ORIGIN
```

```
  Query Match             9.8%;  Score 50.4;  DB 14;  Length 228279;
  Best Local Similarity   46.6%;  Pred. No. 2.5;
  Matches  162;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy         16 TTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAA 75
              |||| ||   |    |||  | |  | || |  | ||  |  | |   ||  || | |||
Db     208568 TTAAAATGGATTAAAAAAATAAATAAATAAAAACTGCTTTTATTCTAGCCAAAGTAAAAC 208627

Qy         76 GCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAA 135
              | |  |  || | | |  ||    | || || | | |     ||  ||     |
Db     208628 AAATAAGTCTTTCTCCAGAATAAAAAACATTATCTGACATACTGTGAAAATGTGCTCTGT 208687

Qy        136 TTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGG 195
              |  |          | | |  |||||| || |     |||||| |||| | ||
Db     208688 TAAACATAATTTGGGAAATATTTAAAAAAGAAAAGAAAATCCAAAGGGGGGACTAATAAT 208747

Qy        196 GTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCT 255
               ||     ||||||  |  |  | | ||||||  |||    ||  | |  | |
Db     208748 TCTGAATTCAACTGCATGCATACACATATATATATATATATATATATATATATATATA 208807

Qy        256 TATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCA 315
              ||| | || |||  |      |||||  | ||  |    || |  |||| | ||      |
Db     208808 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 208867

Qy        316 TGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATA 363
              | | ||  | || | |   ||||||  ||   ||   |  | ||| ||
Db     208868 TATATATATATATATATATATATATATATATATATATATATATATATGTA 208915
```

Search completed: May 11, 2006, 19:14:25
Job time : 1858.78 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 07:09:59 ; Search time 234.661 Seconds
(without alignments)
14626.705 Million cell updates/sec

Title: US-09-423-093B-3_COPY_3361_3875
Perfect score: 515
Sequence: 1 atgtttgtctatagtttaag.....gctggtagtgaaatttgtga 515

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	515	100.0	12441	2	AAX06750	Aax06750 Salmonell
	2	48.8	9.5	8711	4	AAS46700	Aas46700 Tumour su
	3	47.6	9.2	11422	6	ABK39936	Abk39936 Human che
	4	47.6	9.2	11422	6	ABL32218	Abl32218 Human imm
	5	47.2	9.2	17967	6	ABL33014	Abl33014 Human imm
	6	46.8	9.1	6056	6	ABL33026	Abl33026 Human imm
	7	46.4	9.0	19634	8	ABZ10015	Abz10015 Haematopo
	8	46.4	9.0	19634	8	ABZ10161	Abz10161 Haematopo
	9	46.4	9.0	19634	13	ADS89671	Ads89671 Oligonucl
	10	46.4	9.0	19634	13	ADS89397	Ads89397 Oligonucl
	11	46.4	9.0	19659	6	ABL32767	Abl32767 Human imm
	12	46	8.9	15732	4	AAS45388	Aas45388 Chemicall
	13	46	8.9	15732	6	ABK28233	Abk28233 DNA trans
	14	46	8.9	17137	6	ABL32191	Abl32191 Human imm
	15	45	8.7	6087	6	ABL32418	Abl32418 Human imm
	16	44.8	8.7	2000	11	ACL35887	Acl35887 Rice stre
c	17	44.6	8.7	2000	11	ACL37108	Acl37108 Rice stre
c	18	44.6	8.7	4660	14	ADZ70902	Adz70902 Human mat
	19	44.2	8.6	13326	6	ABL33712	Abl33712 Human imm
	20	44	8.5	3927	3	AAA70101	Aaa70101 Plasmodiu
	21	44	8.5	8711	4	AAS46699	Aas46699 Tumour su
	22	43.6	8.5	11047	4	AAS45480	Aas45480 Chemicall
	23	43.6	8.5	11047	6	ABL33985	Abl33985 Human imm
	24	43.6	8.5	11047	6	ABK28402	Abk28402 DNA trans
c	25	43.4	8.4	2000	11	ACL35363	Acl35363 Rice stre
	26	43.4	8.4	6668	4	AAS46418	Aas46418 Tumour su
	27	43.4	8.4	6668	6	ABL33219	Abl33219 Human imm
	28	43.4	8.4	6668	6	ABN80163	Abn80163 Human che
	29	43.4	8.4	6775	6	ABQ67160	Abq67160 Human ang
	30	43.4	8.4	8900	13	ADS89686	Ads89686 Oligonucl
	31	43.4	8.4	8900	13	ADS89412	Ads89412 Oligonucl
	32	43.4	8.4	46951	10	ADE13891	Ade13891 Human opt
	33	43.4	8.4	110000	2	AAV21209_04	Continuation (5 of
	34	43.2	8.4	8453	6	ABN80024	Abn80024 Human che
	35	43.2	8.4	10329	6	ABL34122	Abl34122 Human imm
	36	43	8.3	8237	4	AAS46801	Aas46801 Tumour su
	37	43	8.3	14919	4	AAS46506	Aas46506 Tumour su
c	38	42.8	8.3	2486	3	AAA97037	Aaa97037 Nucleotid
	39	42.8	8.3	6207	4	AAS46777	Aas46777 Tumour su
	40	42.6	8.3	6118	6	ABL33030	Abl33030 Human imm
c	41	42.6	8.3	110000	6	ABA92787_4	Continuation (5 of
	42	42.6	8.3	113515	6	ABL34174	Abl34174 Human imm
c	43	42.4	8.2	1350	5	AAS72599	Aas72599 DNA encod
	44	42.4	8.2	6641	6	ABL32315	Abl32315 Human imm
	45	42.4	8.2	6641	6	ABL54336	Abl54336 Chemicall

ALIGNMENTS

RESULT 1
AAX06750
ID AAX06750 standard; DNA; 12441 BP.
XX
AC AAX06750;
XX
DT 26-APR-1999 (first entry)
XX
DE Salmonella enterica serovar C2 O antigen gene cluster.
XX
KW O antigen; diagnosis; wbaR gene; wzx gene; wbaL gene; wbaQ gene;
KW wxy gene; wbaW gene; wbaZ gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Salmonella enterica.
XX
FH Key Location/Qualifiers
FT CDS 138. .1019
FT /*tag= a
FT /note= "abe gene; encodes AAW88324"
FT CDS 1019. .2359
FT /*tag= b
FT /product= "flippase"
FT /note= "wzx gene, this region is specifically claimed in
FT Claim 9; encodes AAW88325"
FT CDS 2362. .3314
FT /*tag= c
FT /product= "abequosyl_transferase"
FT /note= "wbaR gene, this region is specifically claimed in
FT Claim 9; encodes AAW88326"
FT CDS 3361. .3975
FT /*tag= d
FT /product= "acetyl_transferase"
FT /note= "wbaL gene, this region is specifically claimed in
FT Claim 9; encodes AAW88327"
FT CDS 3977. .5020
FT /*tag= e
FT /product= "rhamnosyl_transferase"
FT /note= "wbaQ gene, this region is specifically claimed in
FT Claim 9; encodes AAW88328"
FT CDS 5114. .6313
FT /*tag= f

```
FT                   /product= "O_antigen_polymerase"
FT                   /note= "wzy gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88329"
FT   CDS             6313. .7323
FT                   /*tag= g
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88330"
FT   CDS             7310. .8467
FT                   /*tag= h
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88331"
FT   CDS             8482. .9903
FT                   /*tag= i
FT                   /note= "manC gene; encodes AAW88332"
FT   CDS             9911. .11347
FT                   /*tag= j
FT                   /note= "manB gene; encodes AAW88333"
FT   CDS             11426. .12441
FT                   /*tag= k
FT                   /note= "wbaP gene; encodes AAW88334"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88324, AAW88325, AAW88326, AAW88327, AAW88328, AAW88329,
DR   AAW88330, AAW88331, AAW88332, AAW88333, AAW88334.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 9; Fig 9; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Salmonella enterica serotype C2 (strain M67) O
CC   antigen. 11 Open reading frames within the gene cluster were identified
CC   on the basis of homology to known sequences. The genes encode proteins
CC   (see AAW88323-34) involved the synthesis of sugars present in the
CC   polysaccharide antigen, and in the transport or processing of
CC   polysaccharide or oligosaccharide units. The use of nucleic acid
CC   molecules derived from assembly and transport genes, particularly wbd
CC   (transferase), wzx (flippase) and wzy (polymerase) genes, within O
CC   antigen gene clusters improves the specificity of methods for the
CC   detection and identification of O antigens, e.g. in testing food- or
CC   faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of S. enterica
XX
SQ   Sequence 12441 BP; 3887 A; 1876 C; 2518 G; 4160 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 515;  DB 2;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 3.3e-113;
  Matches  515;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 3420

Qy         61 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 3480

Qy        121 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 3540

Qy        181 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 3600

Qy        241 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 3660

Qy        301 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT 3720

Qy        361 ATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 ATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA 3780

Qy        421 AATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 AATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGAT 3840

Qy        481 AATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 515
              |||||||||||||||||||||||||||||||||||
Db       3841 AATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 3875


RESULT 2
AAS46700
ID   AAS46700 standard; DNA; 8711 BP.
XX
AC   AAS46700;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #423.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 423; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
```

CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 8711 BP; 2618 A; 83 C; 1763 G; 4247 T; 0 U; 0 Other;

Query Match 9.5%; Score 48.8; DB 4; Length 8711;
Best Local Similarity 48.5%; Pred. No. 0.1;
Matches 165; Conservative 0; Mismatches 172; Indels 3; Gaps 1;

Qy 35 TTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGC 94
Db 2643 TTGTAATTATAAAATTTGGTTTTTGTTGGGAGAAAATAGAAGTGAGGATAAATGTTTATA 2702

Qy 95 TTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAATGTCAAACCTTACT 154
Db 2703 TTGGATTATTAATTTTTAAATTATTAAGATTTGATAATATTTTTGTAGGGAAAGATTATA 2762

Qy 155 CTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGTGATAACTGCTGGA 214
Db 2763 TTGTTATATATAGGTTATATTTTAATTTAGGGATGATGAAGTTTGATATGAATATATGTA 2822

Qy 215 TTGAAGCTGTATATAATTATGGTGATGAAAAATTTGA---ACCTTATTTGTACATAGGTG 271
Db 2823 TATAAATTTGTAATTTGTGTGGTAATTATAAATTTTAGTGATTTGGTTTGTTTAATGATG 2882

Qy 272 ATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAA 331
Db 2883 AATGTAGTGTTATGTTAAATATATAATAAACGATTTTTTTATTAAGTTTATTAAAAGAAT 2942

Qy 332 ACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCA 371
Db 2943 TTTTTTTTTTTTTGGGATGGAGTTTTGTTTTGTCGTTTA 2982

RESULT 3
ABK39936
ID ABK39936 standard; DNA; 11422 BP.
XX
AC ABK39936;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #9 strand 1.
XX
KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO200202806-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007470.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 17; 24pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 11422 BP; 3867 A; 59 C; 1786 G; 5710 T; 0 U; 0 Other;

Query Match 9.2%; Score 47.6; DB 6; Length 11422;
Best Local Similarity 48.0%; Pred. No. 0.21;
Matches 169; Conservative 0; Mismatches 179; Indels 4; Gaps 1;

Qy 24 AAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAAGTT 83
Db 3873 AATAATATAGATTAAAATTTTGTTTTTTAATAGTTATTAGTGGTAAGAAATATAGTAAGT 3932

Qy 84 TCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAATGT 143
Db 3933 TTTTGATTTGTAGAAATTTTAGTTTTTTTATTTGATGAGGTGGTTAATAATATTAATTTT 3992

Qy 144 CAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGT-- 201
Db 3993 TATATTTGTTGAAATAATATAAAATATTTAATATAAATTGTTAATTATAATGATTGGTAT 4052

Qy 202 --GATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATT 259
Db 4053 AAGGTAGAATTATAATAAATGTTAATTATATTTGTTTGTTTGAAAATTTAATATTTTATT 4112

Qy 260 TGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTT 319
Db 4113 AAAAGGTAAAAAAGTGTTTATTTGTAAATATTAATATATGTTTTTTGTTTAGTTTATAGA 4172

Qy 320 TAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCA 371
Db 4173 AAGTTTATAAGTAAGAGATTTAAAAGTGAAAGTTATTATTTATAGAAGATTA 4224

RESULT 4
ABL32218
ID ABL32218 standard; DNA; 11422 BP.

XX
AC ABL32218;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 191.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 191; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 11422 BP; 3867 A; 59 C; 1786 G; 5710 T; 0 U; 0 Other;

Query Match 9.2%; Score 47.6; DB 6; Length 11422;
Best Local Similarity 48.0%; Pred. No. 0.21;
Matches 169; Conservative 0; Mismatches 179; Indels 4; Gaps 1;

```
Qy        24 AAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAAGTT 83
Db      3873 AATAATATAGATTAAAATTTTGTTTTTTAATAGTTATTAGTGGTAAGAAATATAGTAAGT 3932

Qy        84 TCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAATGT 143
Db      3933 TTTTGATTTGTAGAAATTTTAGTTTTTTTATTTGATGAGGTGGTTAATAATATTAATTTT 3992

Qy       144 CAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGT-- 201
Db      3993 TATATTTGTTGAAATAATATAAAATATTTAATATAAATTGTTAATTATAATGATTGGTAT 4052

Qy       202 --GATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATT 259
Db      4053 AAGGTAGAATTATAATAAATGTTAATTATATTTGTTTGTTTGAAAATTTAATATTTTATT 4112

Qy       260 TGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTT 319
Db      4113 AAAAGGTAAAAAAGTGTTTATTTGTAAATATTAATATATGTTTTTTGTTTAGTTTATAGA 4172

Qy       320 TAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCA 371
Db      4173 AAGTTTATAAGTAAGAGATTTAAAAGTGAAAGTTATTATTTATAGAAGATTA 4224
```

RESULT 5
ABL33014
ID ABL33014 standard; DNA; 17967 BP.
XX
AC ABL33014;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 987.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 987; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 17967 BP; 5375 A; 247 C; 3878 G; 8467 T; 0 U; 0 Other;

Query Match 9.2%; Score 47.2; DB 6; Length 17967;
Best Local Similarity 49.6%; Pred. No. 0.29;
Matches 121; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
Qy        83 TTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAATG 142
Db     13531 TTTTTGTTGTTTTTTAAGTGTTTTTAATTTAAATAGTTAATATGTTAGAAGAGTATATTT 13590

Qy       143 TCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGTG 202
Db     13591 TAATATTTTTATTAATATAATAAAAAATTTAATTATAATGTAGTAGAATAAGATTTAATG 13650

Qy       203 ATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGT 262
```

```
Db      13651 TAAGTAGTTTTAATGTAGTTGTATTAAATTAAATTAGAAGTAAAATAGATGAATATATAT 13710

Qy        263 ACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAA 322
              | |||  | |   | ||| |   |  |  ||||| ||  |  ||    ||  |||||
Db      13711 ATATAATTTAAGTTTTATTTAGGATAGTATATGTTTATTATGTTTATAAATTGTGTTTTT 13770

Qy        323 TTTT 326
              |||
Db      13771 GTTT 13774


RESULT 6
ABL33026
ID   ABL33026 standard; DNA; 6056 BP.
XX
AC   ABL33026;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 999.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 999; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6056 BP; 1780 A; 54 C; 1079 G; 3143 T; 0 U; 0 Other;

  Query Match             9.1%;  Score 46.8;  DB 6;  Length 6056;
  Best Local Similarity   47.5%;  Pred. No. 0.29;
  Matches  172;  Conservative    0;  Mismatches  187;  Indels    3;  Gaps    1;

Qy          6 TGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCG 65
              | | |  |  | || |||  | ||   | |||  |||     ||| |  |||  |||
Db       3749 TTTTTCGAAATGAAAATTTTATTTTGTTTTTAAAATAAATAAATTAATAAAAAGTAGTGT 3808

Qy         66 GAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAA 125
                |   ||       |||| ||| | |    |     | || ||||| | |  ||  | |
Db       3809 TTATGTAAGTATTTTGTTTTTTGATTTTAGGAATTTTTTTGTTTTTATATTAATTTTGTA 3868

Qy        126 GAATTTTAAATTGAATGTCAAACCTTACTCTGCAAA---AAATAACACCTCTTCCAAATG 182
                |  ||  | ||  | |||  ||| | ||  ||      |   |        | ||
Db       3869 TTGGTAAAATATAAAAATAAAATGTTATTTTGTTAATTTTTTTTTTTAGTAAAGTAATATA 3928

Qy        183 GGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGA 242
              |  ||||||  |  |  |  | || || |  |    |   ||| | || | ||  |  | |
Db       3929 GTTTAGTATATGTTTAAGAAAAAAATGGTTAAGAAATTTTGTTTTTATTAATTATTTTTA 3988

Qy        243 AAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATAT 302
              | | | |||   |||| | |    |  | | |||     ||||   ||| || ||   ||
Db       3989 AGATTGTGATATTTATATTTGTTTTTTAAACGTTATATTTTATAAGATTATTTTTTAAT 4048

Qy        303 TTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATAT 362
              | |||  ||| |  ||||| | ||   |   || ||   |  ||   |  ||    |
Db       4049 ATGTTGAATATAAGTTTTAAATATATGTATTTTTAAATTGTATTTTAAATATTTTAGGTT 4108

Qy        363 AG 364
              ||
Db       4109 AG 4110


RESULT 7
ABZ10015
ID   ABZ10015 standard; DNA; 19634 BP.
XX
AC   ABZ10015;
XX
DT   16-JAN-2003  (first entry)
XX
DE   Haematopoietic cell proliferation disorder related DNA sequence #155.
XX
KW   Human; haematopoietic cell proliferation disorder; cytostatic;
KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW   cytosine methylation state; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277272-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-EP003401.
XX
PR   26-MAR-2001; 2001US-0278333P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Berlin K,  Braun A,  Distler J,  Guetig D,  Howe A,  Mueller J;
PI   Olek A,  Piepenbrock C,  Adorjan P,  Grabs G,  Lesche R,  Leu E;
PI   Lewin A,  Lipscher E,  Maier S,  Model F,  Mueller V,  Otto T,  Pelet C;
PI   Schwope I,  Ziebarth H;
XX
DR   WPI; 2003-018942/01.
XX
PT   Detecting and differentiating between hematopoietic cell proliferative
PT   disorders, comprises contacting a target nucleic acid with a reagent that
PT   distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS   Claim 28; SEQ ID NO 155; 117pp; English.
XX
CC   The present invention describes a method for detecting and
CC   differentiating between haematopoietic cell proliferative disorders
CC   associated with at least 1 gene and/or their regulatory regions in a
CC   subject. The method comprises contacting a target nucleic acid in a
CC   biological sample obtained from the subject with at least 1 reagent,
CC   which distinguishes between methylated and non-methylated CpG
CC   dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC   represent specifically claimed nucleotide sequences from the present
CC   invention. Oligonucleotides from the present invention can be used: for
CC   differentiating between healthy haematopoietic cells and proliferative
```

CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 19634 BP; 5528 A; 272 C; 3979 G; 9855 T; 0 U; 0 Other;

Query Match 9.0%; Score 46.4; DB 8; Length 19634;
Best Local Similarity 48.4%; Pred. No. 0.46;
Matches 187; Conservative 0; Mismatches 196; Indels 3; Gaps 2;

```
Qy          4 TTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGG 63
              ||||| | |||||| || ||| | |||  |    | |||| |||| | || || | |
Db      11021 TTTGTTTTTAGTTTGAGTTTATAGTTATGTAAATTAATATTATTAGTTAGAAAGGGAAAT 11080

Qy         64 CGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGG 123
               || ||  |   |  |   || ||      |||   ||  |       ||    |||
Db      11081 AGGTAAAAATGTGTTATTGTTTTTTGGAAATTTATTAGAATTTTAAGGTAAGGAATTGAA 11140

Qy        124 AAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGG 183
              |       | | || ||  |  |     | | |   || ||||  |    ||   ||| |
Db      11141 ATTTTAGGTTATTTTAAGATTGATTAGGAATATTATAATAATATTAAAGATTGAGAATTG 11200

Qy        184 GGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAA 243
              | | ||  |   |  ||  || | |      |  ||    | | ||||||| | ||  ||
Db      11201 GATTGTGGGTTTGGGGGAAAT-ATTATATTTTGAAATGATTTTGTAATTATTGAGAATAA 11259

Qy        244 AAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATA-- 301
              | ||   |  |||   |  |||     |     ||||||  | ||||||  ||
Db      11260 ATATGGAATTTTTAGAATTTAATATAGATGGTTTATATTTTAAGATAAAATGTTGTTAAT 11319

Qy        302 TTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATA 361
              ||| ||   | |   ||| |||||| |||   |  ||  |||| ||| ||   | |||
Db      11320 TTTGTTTTATTTATATTTTTATTTTAATAAATATTTGATAGATTGATAGGAATTATAATA 11379

Qy        362 TAGGCGATCATAGCCATGGCAGTTAT 387
              |    | |  |||  |    | || |
Db      11380 TGTTAGGTTGTAGTTAAATGAATTTT 11405
```

RESULT 8
ABZ10161
ID ABZ10161 standard; DNA; 19634 BP.
XX
AC ABZ10161;
XX
DT 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #301.
XX
KW Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 301; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 19634 BP; 5528 A; 0 C; 3979 G; 10127 T; 0 U; 0 Other;

Query Match 9.0%; Score 46.4; DB 8; Length 19634;
Best Local Similarity 48.4%; Pred. No. 0.46;
Matches 187; Conservative 0; Mismatches 196; Indels 3; Gaps 2;

```
Qy          4 TTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGG 63
              ||||| | |||||| || ||| | |||  |    | |||| |||| | || || | |
Db      11021 TTTGTTTTTAGTTTGAGTTTATAGTTATGTAAATTAATATTATTAGTTAGAAAGGGAAAT 11080

Qy         64 CGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGG 123
               || ||  |   |  |   || ||      |||   ||  |       ||    |||
Db      11081 AGGTAAAAATGTGTTATTGTTTTTTGGAAATTTATTAGAATTTTAAGGTAAGGAATTGAA 11140

Qy        124 AAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGG 183
              |       | | || ||  |  |     | | |   || ||||  |    ||   ||| |
Db      11141 ATTTTAGGTTATTTTAAGATTGATTAGGAATATTATAATAATATTAAAGATTGAGAATTG 11200

Qy        184 GGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAA 243
              | | ||  |   |  ||  || | |      |  ||    | | ||||||| | ||  ||
Db      11201 GATTGTGGGTTTGGGGGAAAT-ATTATATTTTGAAATGATTTTGTAATTATTGAGAATAA 11259

Qy        244 AAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATA-- 301
              | ||   |  |||   |  |||     |     ||||||  | ||||||  ||
Db      11260 ATATGGAATTTTTAGAATTTAATATAGATGGTTTATATTTTAAGATAAAATGTTGTTAAT 11319

Qy        302 TTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATA 361
              ||| ||   | |   ||| |||||| |||   |  ||  |||| ||| ||   | |||
Db      11320 TTTGTTTTATTTATATTTTTATTTTAATAAATATTTGATAGATTGATAGGAATTATAATA 11379

Qy        362 TAGGCGATCATAGCCATGGCAGTTAT 387
              |    | |  |||  |    | || |
Db      11380 TGTTAGGTTGTAGTTAAATGAATTTT 11405
```

RESULT 9

```
ADS89671
ID   ADS89671 standard; DNA; 19634 BP.
XX
AC   ADS89671;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Oligonucleotide of the invention SEQ ID NO:687.
XX
KW   ss; cell proliferative disorder; breast; methylation; cytostatic;
KW   gene therapy; single nucleotide polymorphism; SNP.
XX
OS   Unidentified.
XX
PN   WO2004035803-A2.
XX
PD   29-APR-2004.
XX
PF   01-OCT-2003; 2003WO-EP010881.
XX
PR   01-OCT-2002; 2002DE-01045779.
PR   07-JAN-2003; 2003DE-01000096.
PR   17-APR-2003; 2003DE-01017955.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Foekens J,  Harbeck N,  Koenig T,  Maier S,  Martens J,  Model F;
PI   Nimmrich I,  Rujan T,  Schmitt A,  Schmitt M,  Look MP,  Marx A;
XX
DR   WPI; 2004-348468/32.
XX
PT   Predicting responsiveness of a subject with breast cell proliferative
PT   disorder, useful for treating or differentiating breast cell
PT   proliferative disorders comprises analyzing methylation pattern of a
PT   genomic DNA from the subject.
XX
PS   Disclosure; SEQ ID NO 687; 104pp; English.
XX
CC   The invention relates to a novel method for predicting the responsiveness
CC   of a subject with a cell proliferative disorder of the breast tissues to
CC   a therapy comprising analysing the methylation pattern of a target
CC   nucleic acid by contacting at least one of the target nucleic acids in a
CC   biological sample obtained from the subject prior to or during treatment.
CC   The method of the invention has cytostatic activity, and may have a use
CC   in gene therapy. The set of oligonucleotides comprising at least two of
CC   the oligomers are useful for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms (SNPs) within the sequences. The
CC   methods, nucleic acid, oligonucleotide, and kit are useful for the
CC   treatment, characterisation, classification and/or differentiation, of
CC   breast cell proliferative disorders. The method is also useful for
CC   predicting the responsiveness of a subject with a cell proliferative
CC   disorder of the breast tissues to a therapy. The present sequence is used
CC   in the exemplification of the invention.
XX
SQ   Sequence 19634 BP; 5528 A; 0 C; 3979 G; 10127 T; 0 U; 0 Other;

  Query Match             9.0%;  Score 46.4;  DB 13;  Length 19634;
  Best Local Similarity   48.4%;  Pred. No. 0.46;
  Matches  187;  Conservative    0;  Mismatches  196;  Indels    3;  Gaps    2;

Qy          4 TTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGG 63
              |||||| | |||||| || ||| | |||  |     | |||| |||| | || || |  |
Db      11021 TTTGTTTTTAGTTTGAGTTTATAGTTATGTAAATTAATATTATTAGTTAGAAAGGGAAAT 11080

Qy         64 CGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGG 123
               || ||  |   |  |   || ||       |||   ||  |       ||    |||
Db      11081 AGGTAAAAATGTGTTATTGTTTTTTGGAAATTTATTAGAATTTTAAGGTAAGGAATTGAA 11140

Qy        124 AAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGG 183
              |       | | || ||  |  |     | | |   || ||||  |    ||   ||| |
Db      11141 ATTTTAGGTTATTTTAAGATTGATTAGGAATATTATAATAATATTAAAGATTGAGAATTG 11200

Qy        184 GGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAA 243
              |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      11201 GATTGTGGGTTTGGGGGAAAT-ATTATATTTTGAAATGATTTTGTAATTATTGAGAATAA 11259

Qy        244 AAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATA-- 301
              | ||   |  |||    |  |||          |      ||||||  | ||||||  ||
Db      11260 ATATGGAATTTTTAGAATTTAATATAGATGGTTTATATTTTAAGATAAAATGTTGTTAAT 11319

Qy        302 TTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATA 361
              ||| ||   | |    ||| ||||||  |||    |  ||  |||| ||| ||   | |||
Db      11320 TTTGTTTTATTTATATTTTTATTTTAATAAATATTTGATAGATTGATAGGAATTATAATA 11379

Qy        362 TAGGCGATCATAGCCATGGCAGTTAT 387
              |    | |  |||  |    | || |
Db      11380 TGTTAGGTTGTAGTTAAATGAATTTT 11405


RESULT 10
ADS89397
ID   ADS89397 standard; DNA; 19634 BP.
XX
AC   ADS89397;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Oligonucleotide of the invention SEQ ID NO:413.
XX
KW   ss; cell proliferative disorder; breast; methylation; cytostatic;
KW   gene therapy; single nucleotide polymorphism; SNP.
XX
OS   Unidentified.
XX
PN   WO2004035803-A2.
XX
PD   29-APR-2004.
XX
PF   01-OCT-2003; 2003WO-EP010881.
XX
PR   01-OCT-2002; 2002DE-01045779.
PR   07-JAN-2003; 2003DE-01000096.
PR   17-APR-2003; 2003DE-01017955.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Foekens J,  Harbeck N,  Koenig T,  Maier S,  Martens J,  Model F;
PI   Nimmrich I,  Rujan T,  Schmitt A,  Schmitt M,  Look MP,  Marx A;
XX
DR   WPI; 2004-348468/32.
XX
PT   Predicting responsiveness of a subject with breast cell proliferative
PT   disorder, useful for treating or differentiating breast cell
PT   proliferative disorders comprises analyzing methylation pattern of a
PT   genomic DNA from the subject.
XX
PS   Disclosure; SEQ ID NO 413; 104pp; English.
XX
CC   The invention relates to a novel method for predicting the responsiveness
CC   of a subject with a cell proliferative disorder of the breast tissues to
CC   a therapy comprising analysing the methylation pattern of a target
CC   nucleic acid by contacting at least one of the target nucleic acids in a
CC   biological sample obtained from the subject prior to or during treatment.
CC   The method of the invention has cytostatic activity, and may have a use
CC   in gene therapy. The set of oligonucleotides comprising at least two of
CC   the oligomers are useful for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms (SNPs) within the sequences. The
CC   methods, nucleic acid, oligonucleotide, and kit are useful for the
CC   treatment, characterisation, classification and/or differentiation, of
CC   breast cell proliferative disorders. The method is also useful for
CC   predicting the responsiveness of a subject with a cell proliferative
CC   disorder of the breast tissues to a therapy. The present sequence is used
CC   in the exemplification of the invention.
XX
SQ   Sequence 19634 BP; 5528 A; 272 C; 3979 G; 9855 T; 0 U; 0 Other;
```

```
  Query Match             9.0%;  Score 46.4;  DB 13;  Length 19634;
  Best Local Similarity   48.4%;  Pred. No. 0.46;
  Matches  187;  Conservative    0;  Mismatches  196;  Indels    3;  Gaps    2;

Qy          4 TTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGG 63
              ||||| | |||||| || ||| | |||   |     | |||| |||| | || || | |
Db      11021 TTTGTTTTTAGTTTGAGTTTATAGTTATGTAAATTAATATTATTAGTTAGAAAGGGAAAT 11080

Qy         64 CGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGG 123
               || ||  |   |  |  || ||      |||   ||  |       ||    |||
Db      11081 AGGTAAAAATGTGTTATTGTTTTTTGGAAATTTATTAGAATTTTAAGGTAAGGAATTGAA 11140

Qy        124 AAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGG 183
              |       | | || ||  |  |     | | |   || ||||  |    ||   ||| |
Db      11141 ATTTTAGGTTATTTTAAGATTGATTAGGAATATTATAATAATATTAAAGATTGAGAATTG 11200

Qy        184 GGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAA 243
              | | ||  |   |  ||  || | |       |  ||    | | ||||||| | ||  ||
Db      11201 GATTGTGGGTTTGGGGGAAAT-ATTATATTTTGAAATGATTTTGTAATTATTGAGAATAA 11259

Qy        244 AAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATA-- 301
              | ||   |  |||     |  |||       |     ||||||  | ||||||  ||
Db      11260 ATATGGAATTTTTAGAATTTAATATAGATGGTTTATATTTTAAGATAAAATGTTGTTAAT 11319

Qy        302 TTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATA 361
              ||| ||   | |    ||| |||||| |||   |  ||  |||| ||| ||   | |||
Db      11320 TTTGTTTTATTTATATTTTTATTTTAATAAATATTTGATAGATTGATAGGAATTATAATA 11379

Qy        362 TAGGCGATCATAGCCATGGCAGTTAT 387
              |    | |  |||  |    | || |
Db      11380 TGTTAGGTTGTAGTTAAATGAATTTT 11405


RESULT 11
ABL32767
ID   ABL32767 standard; DNA; 19659 BP.
XX
AC   ABL32767;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 740.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 740; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 19659 BP; 6577 A; 68 C; 3643 G; 9371 T; 0 U; 0 Other;

  Query Match             9.0%;  Score 46.4;  DB 6;  Length 19659;
  Best Local Similarity   46.8%;  Pred. No. 0.46;
  Matches  178;  Conservative    0;  Mismatches  201;  Indels    1;  Gaps    1;

Qy         20 GATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAA 79
              ||| | |   || |   |   ||      | ||| ||| |||     |     |   |
Db       3238 GATGAGAGAAAAGTGGGAAATTAGATGGAATGAATAAGGTTTAAAGAGGTTTTGAAATTA 3297

Qy         80 AGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGA 139
               ||| |||   |  ||||         ||  ||  | |   |||| ||| |||||   |
Db       3298 TTTTTGTTGGAATATTTAGTTTTGGGTATAATAGGAAGTAAGGGAGGAAGTTTAAGAGGG 3357

Qy        140 ATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTG 199
              || |  |       |    ||||||  |          | || |||    |  ||
Db       3358 ATATTGAGAAAGTTTAGAATTAAAATATAAAAATAATTTGTAGAGGAAGTTGTTGTATTA 3417

Qy        200 GTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATT 259
                |   |    || | |      || || || || | |   ||  ||  ||  |||||
Db       3418 TGATTTTTTAGGTGAGTTATTTAAAATGTATTTTTGATAAAATAATTTTATAAATTTATT 3477

Qy        260 TGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTT 319
              |||  | ||  |      |  |||||  ||||| ||| || | | ||   |||  | ||
Db       3478 TGTTTAAAGAGAAGATATTTAGTTTATAAGATAAAGTTAATGTATGTTAATTATTGT-TT 3536

Qy        320 TAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATG 379
              |||||||  || |   || || |||     |    | ||| ||  |    |  ||   ||
Db       3537 TAATTTTTAAATATAGTAAATAAATATTAAAGTTAATTTGTAATGTGATTATTATTAATT 3596

Qy        380 GCAGTTATAAAGTATGCAGT 399
              |   || ||   |||| | |
Db       3597 GATTTTTTATTTTATGGAAT 3616


RESULT 12
AAS45388
ID   AAS45388 standard; DNA; 15732 BP.
XX
AC   AAS45388;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Chemically pretreated genomic DNA associated with cell cycle #47.
XX
KW   Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW   human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW   graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW   arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW   immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168911-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002945.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
```

PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602751/68.
XX
PT Designing primers and probes for analyzing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle.
XX
PS Claim 1; SEQ ID NO 93; 28pp; English.
XX
CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers
XX
SQ Sequence 15732 BP; 4638 A; 70 C; 2672 G; 8352 T; 0 U; 0 Other;

Query Match 8.9%; Score 46; DB 4; Length 15732;
Best Local Similarity 47.7%; Pred. No. 0.55;
Matches 168; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

Qy 22 TTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAAG 81
Db 13292 TTTTGATTACGTTATATTATTTAAGGGAGAAAGTTAAATAGGGATAAAGAAAAATATTTA 13351

Qy 82 TTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAAT 141
Db 13352 GAAGTTTTTTATTATTTTGAATGTGGTTTAATTTTGGTTTGGAATTTTTTTTGTTGTAAT 13411

Qy 142 GTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGT 201
Db 13412 TAATTTTTAATAGTTGAAAATATTAGTTTGGTTTAAAAATTGTTTTTTATTTGGTGTTTT 13471

Qy 202 GATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTG 261
Db 13472 TA----TGGTATAATGAAAGTTTGGAACGTTTTTGTTTGTAATTTTGTTAATGTTATTTT 13527

Qy 262 TACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTA 321
Db 13528 TAGTGAGAAGAGAATAATTTTTTAAATTATAATATTTAAATAATATTTTGATTAGATATA 13587

Qy 322 ATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATA 373
Db 13588 AAATTTTAGTTTGAAAAATATTTTTGTTTGAAAATTGATTTATAGTATTTTA 13639

RESULT 13
ABK28233
ID ABK28233 standard; DNA; 15732 BP.
XX
AC ABK28233;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA transcription associated genomic DNA #54.
XX
KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX
OS Unidentified.
XX
PN WO200192565-A2.
XX
PD 06-DEC-2001.
XX
PF 06-APR-2001; 2001WO-EP003973.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-090046/12.
XX
PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological disorders,
PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT cancer.
XX
PS Claim 1; SEQ ID NO 107; 32pp; English.
XX
CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office
XX
SQ Sequence 15732 BP; 4638 A; 70 C; 2672 G; 8352 T; 0 U; 0 Other;

Query Match 8.9%; Score 46; DB 6; Length 15732;
Best Local Similarity 47.7%; Pred. No. 0.55;
Matches 168; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

Qy 22 TTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAAG 81
Db 13292 TTTTGATTACGTTATATTATTTAAGGGAGAAAGTTAAATAGGGATAAAGAAAAATATTTA 13351

Qy 82 TTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAAT 141
Db 13352 GAAGTTTTTTATTATTTTGAATGTGGTTTAATTTTGGTTTGGAATTTTTTTTGTTGTAAT 13411

```
Qy        142 GTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGT 201
                  || ||| | ||      ||  |||| |  |  |||  ||  |  |
Db      13412 TAATTTTTAATAGTTGAAAATATTAGTTTGGTTTAAAAATTGTTTTTTATTTGGTGTTTT 13471

Qy        202 GATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTG 261
              |    || |  | ||||  | |  |   || | ||    ||   | | ||  ||||||
Db      13472 TA----TGGTATAATGAAAGTTTGGAACGTTTTTGTTTGTAATTTTGTTAATGTTATTTT 13527

Qy        262 TACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTA 321
              ||   ||  ||    ||  | ||||| | ||||| || | ||    |    || |  | ||
Db      13528 TAGTGAGAAGAGAATAATTTTTTAAATTATAATATTTAAATAATATTTTGATTAGATATA 13587

Qy        322 ATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATA 373
              |  ||  |    || | ||   || ||   ||| || || ||    ||  ||
Db      13588 AAATTTTAGTTTGAAAAATATTTTTGTTTGAAAATTGATTTATAGTATTTTA 13639
```

RESULT 14
ABL32191
ID ABL32191 standard; DNA; 17137 BP.
XX
AC ABL32191;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 164.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 164; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 17137 BP; 5908 A; 240 C; 3151 G; 7838 T; 0 U; 0 Other;

Query Match 8.9%; Score 46; DB 6; Length 17137;
Best Local Similarity 47.0%; Pred. No. 0.56;

Matches 174; Conservative 0; Mismatches 195; Indels 1; Gaps 1;

```
Qy         21 ATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAA 80
               |||| | ||  ||       | |   |||||   |||  ||       |||  | ||
Db       2336 ATTATATTTTTATAAATAGAAAGATTTATTTTTTAATATTGTATTTAAAAAAGAACGTTT 2395

Qy         81 GTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAA 140
               ||| |   |     |       |||||  | |||| |  || |      | | | || | |
Db       2396 ATTTTTATATAATATATAGATATATGCGTATAAATTTATTTGCGTATAATATTAAGTTTA 2455

Qy        141 TGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGG 200
                 |||  ||| | |  |||| ||   |  | ||  ||||  || | |||    |||
Db       2456 ATAAAAATATTATTTTTTAAAATAT-TTATATATTTTAAATATGGAATTATAGTAGTTAA 2514

Qy        201 TGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTT 260
               |||| |    |  |||    |||   | ||| |  | | |||  |  |  |  || |||
Db       2515 TGATTAAGAATTTTTTGTTTTTGTTATTTATTGTTTTTAAGAATTAAATTGAATTTTTTT 2574

Qy        261 GTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTT 320
               |    |   ||  ||  |  ||| ||| | | ||    | | | ||  || |  |
Db       2575 TTGGTAAAAGGAATGTTAAAATTTTAGTAAAATTGGGTTTTTGTTTTATGTTTTTTTGAA 2634

Qy        321 AATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATGG 380
               |||||  |      || |  || |||      |||||  |   ||  ||||  | |
Db       2635 TATTTTGTAGTTGATATTAGTTTTTTGTATTTTTTTTTTATTAATTGGAAAATAGTAAGGT 2694

Qy        381 CAGTTATAAA 390
               |  || | |
Db       2695 TAAATAAATA 2704
```

RESULT 15
ABL32418
ID ABL32418 standard; DNA; 6087 BP.
XX
AC ABL32418;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 391.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 391; 32pp + Sequence Listing; German.
XX

CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6087 BP; 1530 A; 57 C; 1463 G; 3037 T; 0 U; 0 Other;

Query Match 8.7%; Score 45; DB 6; Length 6087;
Best Local Similarity 53.8%; Pred. No. 0.79;
Matches 93; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy        190 ATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTT 249
              ||| || ||  | ||    | | | |||     | || |     |   | |||  ||||
Db       4290 ATGGGGTTTTTTTATGTTGGTTAGGTTGGTTTCGAATTTTTGATTTTAGGTGATTTATTT 4349

Qy        250 GAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGC 309
              |   |   ||| || |  | ||    ||||||| ||||| || |||||    ||   ||
Db       4350 GTTTCGGTTTTTTAAAGTGTTGGAATTATATGTGTAAGTTATTATGTTTGGTTTAAATGT 4409

Qy        310 GTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATAT 362
              ||| ||  ||| ||||   ||   |||||   |||   ||  |   ||| ||
Db       4410 TTATTATTATTATTTTTTATAAGAAATATAGGTATTTTAAGTTATTATTAAAT 4462
```

Search completed: May 11, 2006, 09:42:41
Job time : 237.661 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 09:43:19 ; Search time 1604.57 Seconds
(without alignments)
15016.694 Million cell updates/sec

Title: US-09-423-093B-3_COPY_3361_3875
Perfect score: 515
Sequence: 1 atgtttgtctatagtttaag.....gctggtagtgaaatttgtga 515

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	58.6	11.4	1204	10	CNS016E2	AL106628 Drosophil
2	58.2	11.3	1101	10	CNS0106X	AL098595 Drosophil
c 3	56.4	11.0	1101	10	CNS0039G	AL063921 Drosophil
c 4	56.2	10.9	1101	10	CNS00EVL	AL069706 Drosophil
c 5	55.6	10.8	846	10	CZ861692	CZ861692 OC__Ba025
c 6	55.6	10.8	1101	10	CNS00FVE	AL071298 Drosophil
7	51.8	10.1	1101	10	CNS0100X	AL098379 Drosophil
c 8	51	9.9	1072	10	CL101205	CL101205 ISB1-38C6
9	50.6	9.8	417	1	AW481292	AW481292 35294 MAR
10	50.6	9.8	599	7	CO884183	CO884183 BovGen_12
c 11	50.6	9.8	603	6	CB167163	CB167163 GWM602702
c 12	50.6	9.8	659	1	AM027065	AM027065 AM027065
13	50.6	9.8	662	1	AM026437	AM026437 AM026437
14	50.6	9.8	663	7	CO891008	CO891008 BovGen_19
c 15	50.6	9.8	679	1	AM005285	AM005285 AM005285
c 16	50.6	9.8	679	1	AM005607	AM005607 AM005607
17	50.6	9.8	679	1	AM006104	AM006104 AM006104
18	50.6	9.8	679	1	AM020572	AM020572 AM020572
19	50.6	9.8	685	1	AM006457	AM006457 AM006457
c 20	50.6	9.8	711	6	CB468267	734077 MA
21	50.6	9.8	713	7	CN432776	CN432776 BE030003A
c 22	50.6	9.8	759	7	CK847205	CK847205 969812 MA
c 23	50.6	9.8	847	7	CK776539	CK776539 967858 MA
c 24	50.6	9.8	909	1	AM023833	AM023833 AM023833
25	50.2	9.7	702	9	BH556702	BH556702 BOHPN12TF
26	49.8	9.7	552	10	CZ776361	CZ776361 OC__Ba013
c 27	49.6	9.6	952	10	CNS016RS	AL107122 Drosophil
28	49.4	9.6	750	9	BH549319	BH549319 BOHNH03TR
c 29	49.2	9.6	1056	5	BQ961077	BQ961077 AGENCOURT
c 30	49	9.5	688	1	AM020235	AM020235 AM020235
31	48.8	9.5	1274	1	AJ928756	AJ928756 AJ928756
c 32	48.6	9.4	727	9	BZ037603	BZ037603 oei24a06.
33	48.6	9.4	734	9	BH552664	BH552664 BOHAB33TR
34	48.4	9.4	588	8	DN332129	DN332129 LIB3184-0
35	48.4	9.4	928	10	CNS00DKY	AL071865 Drosophil
36	48.4	9.4	1238	10	CL650442	CL650442 CH213-288
c 37	48.2	9.4	1042	10	CL461494	CL461494 SAIL_1148
38	47.6	9.2	1027	10	CNS02T50	AL212733 Tetraodon
c 39	47.2	9.2	1101	10	CNS003BD	AL064091 Drosophil
c 40	47	9.1	919	10	CNS005RL	AL061409 Drosophil
41	47	9.1	997	10	CNS010RM	AL099340 Drosophil
c 42	47	9.1	1101	10	CW942516	CW942516 TcB24.1_C
43	47	9.1	1101	10	CNS000D1	AL065414 Drosophil
c 44	46.8	9.1	463	10	AG233584	AG233584 Lotus cor
45	46.4	9.0	482	1	AJ692447	AJ692447 AJ692447

ALIGNMENTS

RESULT 1
CNS016E2/c
LOCUS CNS016E2 1204 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106628
VERSION AL106628.1 GI:5622852
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1204)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1204
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN15A12"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"
ORIGIN

Query Match 11.4%; Score 58.6; DB 10; Length 1204;
Best Local Similarity 24.5%; Pred. No. 0.0039;
Matches 124; Conservative 160; Mismatches 220; Indels 3; Gaps 1;

Qy 11 ATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAAT 70

```
Db       1095 WKKKAWAAAAAAAKWAAAKWKATKKDDDDKDKAAADAKKKKKKAKADADKAADKDKDKKAK 1036

Qy         71 CAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATT 130

Db       1035 ADAWDDADDAAKKKKKAKKKDKKKDDATDADWAAKAAKKKDKKAWKKKKDKKKAKKAKAW 976

Qy        131 TTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTA 190

Db        975 KKKDKKDKKDKKKKKKKDKKAKKKKKAAAKDAAAAAAAAAKKKDKKKDAAKAAAKKKKKD 916

Qy        191 TGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATG---GTGATGAAAAAT 247

Db        915 AAKKKKAKKKDDAWAADADKAAAKKKAAAKKKDKKKDAADKKAAKDAAKKKKDKAAKAAD 856

Qy        248 TTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTT 307

Db        855 KDAAKKWDKKKKAGGAARAAAKKGKKGGGRKGGGAGGGRAAAAGATWTTWAKTDTTKTKA 796

Qy        308 GCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCG 367

Db        795 GDAKTTTTKKTTTWATKAGAKWAKKTRAAAKDARGKDAATTAKAARAAGAAATKGGGAAR 736

Qy        368 ATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAAATAGAACCGCCAGCAAATAAGC 427

Db        735 AGGAAGGKDAGARAKGKAAAKAGATKGAAAATACMWAAAGRAWAGWATATGWATRGAAGA 676

Qy        428 CATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGATAATGCAG 487

Db        675 WACGGGWAGAGGGGGGAAAGAYTGACGTBKWTAGDGGTTTTTTKTAGAGATAMGTTMMGG 616

Qy        488 TAATTCTGGCTGGTAGTGAAATTTGTG 514

Db        615 WAATTGAMGGTTGGAATTGGGGTTGTG 589
```

RESULT 2
CNS0106X
LOCUS CNS0106X 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL098595
VERSION AL098595.1 GI:5610206
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN03K20"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"
ORIGIN

Query Match 11.3%; Score 58.2; DB 10; Length 1101;
Best Local Similarity 20.2%; Pred. No. 0.0048;
Matches 96; Conservative 200; Mismatches 177; Indels 2; Gaps 2;

```
Qy          1 ATGTTTGTCTATAGTTTA-AGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGT 59

Db        620 ATGTKTATRTWTAATKTAKRTADTKDTRTAATATAKATKKKARKDKDTAKDAWKWDKATK 679

Qy         60 TAGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGT 119

Db        680 AKAKKKAAKAKAKAAAKTATGKDARKAKAKAKAKATKRAKARAWDTATWTATAADAAADK 739

Qy        120 TGGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAA 179

Db        740 GAKDKAKAKAKADARDKRRWDAKKDKRKRAKAKAAKAADDADADAKDAKAADAKAKADAA 799

Qy        180 ATGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGA 239

Db        800 DADDGRGDGKKKRAKDRKKKKKKDKAWGDKKKAKKDKAAAKAKADAAGDAKAKRRRAGDK 859

Qy        240 TGAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCA 299

Db        860 DKAKADAKAAKKAKAKDDDDAKAATKAKKATKAKKDKAKAKKKKKKKDKDAKAKAKADA 919

Qy        300 TATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTA 359

Db        920 KDKDDKDDDKDDKKDAKADKKKWKKDRAKDKAKKDDDKKDKAKDKDKDKKDADAKAWAKA 979

Qy        360 TATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGA-AAATAGAACCGCCAG 418

Db        980 KADADAKAKAKAKAKAADDDAAADAKAKAADAKAKAKAKADDAAKAKAKADAKAADADAK 1039

Qy        419 CAAATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGAT 473

Db       1040 AKAAKRAAKAKADADAAKAKAKADDKDKDAKKDDADKAKKAKAKDAKAKAKAKAK 1094
```

RESULT 3
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match            11.0%;  Score 56.4;  DB 10;  Length 1101;
  Best Local Similarity  14.3%;  Pred. No. 0.013;
  Matches   61;  Conservative  220;  Mismatches  144;  Indels    3;  Gaps    1;

Qy         65 GGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGA 124
              |:|::: :: :: : ::::: :|::: :  : |: ::: : : : ::: |   :|   :
Db       1101 GKARRWGDDTWDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWADDDGAGTWWTATWW 1042

Qy        125 AGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGG 184
              : :::: |:: ::: :: :  ||  :|| | :: |::|::  : :     :  || : :|
Db       1041 WWWWWWATWDTWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADG 982

Qy        185 GTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAA 244
                :::  : ::::::::: ::: :| :::   :||:   ::::::::: : :::  :::
Db        981 AGRRDGGRKRKDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRK 922

Qy        245 AATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTT 304
              |::::||   ::: : ::| :: :||::: ::::::::::| ||  :| : :   :::::
Db        921 ADDDDGAGDKDDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWW 862

Qy        305 CTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAG 364
              : : :::| :   :::: :: ::|:|:|:  ::|:|: |:| :: :  ::||:    ::::
Db        861 GWADADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRR 802

Qy        365 GCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCAAATA 424
              : ::  :::: :::: | |: |:   |  : ::: : : :|: ||        :::::
Db        801 DRKRADDKRDAADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAA---DRTWDRDDDD 745

Qy        425 AGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGATAATG 484
              ::  | |||:: :  ::: :   :| :::|:|::||:::: : ::||: |: : | :::
Db        744 DRDRAGTAGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKR 685

Qy        485 CAGTAATT 492
               ::::||:
Db        684 RRRDRATW 677


RESULT 4
CNS00EVL/c
LOCUS       CNS00EVL                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match            10.9%;  Score 56.2;  DB 10;  Length 1101;
  Best Local Similarity  34.2%;  Pred. No. 0.014;
  Matches  118;  Conservative   62;  Mismatches  165;  Indels    0;  Gaps    0;

Qy          1 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 60
              ||   | | |:|| | ::| ||  |:|| :||:  ||: :|:: ||:| | : ||:| ||
Db       1012 ATAAATATATWTAWTAWWATATATAWATAWAAWTATAWAWTWWAATWAATWWATAWAWTT 953

Qy         61 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 120
                  :||| :|:   :   || || || |  |:     ||:|  || || ::|
Db        952 TAATAAWAATATWAWWWTWTTATTTTTTTTTTTTTWTTTWTATWTWTTTATATWWTWWAA 893

Qy        121 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 180
                |: ||:|:|| || :|  :|    || | :       :|:     :  |  :|
Db        892 AAWAWWAAWTWTATTTTTWTTATWATTWATATTAWATTTTTTWTWTATTWWTATATWAT 833

Qy        181 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 240
              :   :  |:|     :|  : ||   |  :   |: |   | |:|||::: :  | ::
Db        832 WWTATWTWTWTTTWATWTATWTATTTATWTWTWTTWTATTWTATWTATAWWWTWWATWWW 773

Qy        241 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 300
              |:||:| |  |  :   || |:  :|  :  :  |   | |||  | :    :| |:  |
Db        772 TAWAAWTATATATWWATATTATWWTWATTWTTWTTTTATTWTTTTAAWTTATWTATWATT 713

Qy        301 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATT 345
              :||: |:   | |  | |||  :| |  :|||  || || | |::
Db        712 WTTWATWTTTTTTTTTTTTTTTWTATTTWAAATTATTTAATTAWW 668


RESULT 5
CZ861692/c
LOCUS       CZ861692                846 bp    DNA     linear   GSS 27-JUL-2005
DEFINITION  OC__Ba0255G10.f OC__Ba Oryza coarctata genomic clone OC__Ba0255G10
            5', genomic survey sequence.
ACCESSION   CZ861692
VERSION     CZ861692.1  GI:71336911
KEYWORDS    GSS.
SOURCE      Oryza coarctata (Porteresia coarctata)
  ORGANISM  Oryza coarctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   Kim,H., Collura,K., Wissotski,M., Byrne,M., Stum,D., Smart,D.,
            Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
            Wing,R.
  TITLE     OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
```

```
          Email: rwing@genome.arizona.edu
          PCR PRimers
          FORWARD: TAA TAC GAC TCA CTA TAG GG
          BACKWARD: CAC TCA TTA GGC ACC CCA
          Plate: 0255  row: G  column: 10
          Seq primer: TAA TAC GAC TCA CTA TAG GG
          Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .846
                     /organism="Oryza coarctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:77588"
                     /clone="OC__Ba0255G10"
                     /tissue_type="leaves"
                     /dev_stage="mature"
                     /lab_host="DH10B"
                     /clone_lib="OC__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             10.8%;  Score 55.6;  DB 10;  Length 846;
  Best Local Similarity   48.4%;  Pred. No. 0.019;
  Matches  154;  Conservative    0;  Mismatches  164;  Indels    0;  Gaps    0;

Qy        160 AAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAA 219
              ||||  |  |||   |  ||  |  |      |  ||  | | |  | ||| | |||
Db        804 AAAAGGACGACCCAGTTGAATAGTAGACCTTGATATTTTTTTAAGGAATTTTGGTTGGAA 745

Qy        220 GCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATA 279
              |||  | |  ||  ||| || | |  || ||||||| ||  | ||| | | ||  | |
Db        744 ATGTTATTTGACAATTTTGAAGAGATAATTTAACCTTAATTTGATATATGAGCTCACAAA 685

Qy        280 TGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATA 339
                | | ||  | ||       |   || ||| |||  || |   ||||| ||||  ||| |
Db        684 AATCTTAGGGTTACAAACAAGGCTTATTGTGTAGAATTTGGTATTTTTGAAATGGATTTG 625

Qy        340 TTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGT 399
              |||  |||  ||| ||  || | |  |  | |              |  | || | |
Db        624 AATTTTGATAGAAAATTTGGTAGATGATAGGAAGGGAGGATTTAAAGAAGGATCTGTATT 565

Qy        400 CCGAAAATAGAACCGCCAGCAAATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGT 459
              |  ||  || ||     | |||  |    | | | ||| ||  |   | ||| |   |
Db        564 CAAAATTTAAAATATGGAACAAGGATTTAGTGAAGGGATTTTTGTGGGACTAAGAGTTGG 505

Qy        460 AATTGCTGCTGGATTGGA 477
              |   | |    |||| ||
Db        504 AGGGGATATATGATTTGA 487


RESULT 6
CNS00FVE/c
LOCUS       CNS00FVE                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR32P18 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071298
VERSION     AL071298.1  GI:4951138
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR32P18"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             10.8%;  Score 55.6;  DB 10;  Length 1101;
  Best Local Similarity   28.6%;  Pred. No. 0.019;
  Matches  147;  Conservative  125;  Mismatches  240;  Indels    2;  Gaps    1;

Qy          2 TGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTA 61
              :|  ::: |:|  |::||  || :|::|  :| |: : : :  : |::| :: :   ||
Db       1098 KGGRWKWTTWTTATWWAATKTTGDAWWTKGWTWTDTDGRGRRGWATWWTWDKAWTTTTTK 1039

Qy         62 GGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTG 121
              |    | |  : :  :|: ||: :  |: : :    :  :|  | : |:| : || |
Db       1038 GTTTTTAGTGKWTWTGRAWTTTKGWAATDGDTADRTADRATWAATAAKGTWAGRRGGATA 979

Qy        122 GGAAGAATTTTAA--ATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAA 179
              |:  ||: |:|   :| |:: :  |:  ::| : :|  ::   | :|  |    :|  |
Db        978 AAAWATATKATWAKRGKTTADDTDGKAWWWDWATWWWGARDGKAGWATAAWDGRWTTWAT 919

Qy        180 ATGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGA 239
              | :|  |: |  | |:|       :|| :| |: :: : ||  | |:::| |||: ||
Db        918 AAKGTWTWTTTGGGGKGGGAWGATWAAAKGRTDAWWAKGAGGGGAAWWKATTTAWAGTWT 859

Qy        240 TGAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCA 299
              ||:||:  :  ::  ::|:   :| | | : :|   :| ::::: | ::||| :  |: |
Db        858 TGRAADGAWGWDWGTKDADGRTDADAAAWDADAAWAWAGWWDWWRATRWGATTRATTDAA 799

Qy        300 TATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTA 359
              ||  | |   :  : |: ::|   ::|   :||       | | ||| |:: :  ||  | ||
Db        798 TAGATTTGWWKGTKTAWAWWTKKAWKTTWGWAAWTTATTGTGAAATRRARAAAATWTRTA 739

Qy        360 TATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGC 419
              |   | | | | :      : | :||: |  ::      |   | ||::: |  |  |
Db        738 AAAGAGATAGAARWTAAGATKTATWTADTATWKWATTKWTTTTATAAWWKTAKWGGAAWK 679

Qy        420 AAATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGA 479
              :| |: :  || :  : | ||:  |  :| ||:||:|  :||  |    : | :| |:|
Db        678 WATTWTDTRATAWAWKWAAATWATTTTKAATAWAADATTWAAAGGGATWWAGTKTWGWGK 619

Qy        480 TAATGCAGTAATTCTGGCTGGTAGTGAAATTTGT 513
              : ::: | : ::|  |  |   | |    ||| |
Db        618 WTWKDAAWKTWWTTGGATTAAAATTATGGTTTAT 585


RESULT 7
CNS0100X
LOCUS       CNS0100X                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098379
VERSION     AL098379.1  GI:5609990
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
```

```
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN03G04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             10.1%;  Score 51.8;  DB 10;  Length 1101;
  Best Local Similarity   17.9%;  Pred. No. 0.15;
  Matches   73;  Conservative  171;  Mismatches  163;  Indels    0;  Gaps    0;

Qy          4 TTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGG 63
              :|:|: : :: |:|:   :|: : ||  :: :: :  | : ::: ||  :|: : :|:::
Db        443 DTWGKADTDWWTDTWKTTWTWTWTTTTGWKKWDTKTKTKWKDDDGTTWTDTWKDWKTDKD 502

Qy         64 CGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGG 123
              :: : :::|:   :: :::: : ::: |: :| : : :| |: :|:  :: |: :  |:
Db        503 TDKTWTDVWADNAGWWTDKDKTKAKKKRTKTDTKKTKTRTGTKKKTDRGMRKTDTWGWGD 562

Qy        124 AAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGG 183
              :: ::|:  :|:    |::: |:   :|: | |: :      :  :  : ::  : :::
Db        563 WRADRTDGARADRAANAKDKDADGKADTDATKTDGWGTTTKADTTDDTDADWWDWKDWDW 622

Qy        184 GGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAA 243
              :: : :| : :|  :|::| :   |: ::: |::    |:| : : ::: ::: : :: :
Db        623 KDADTKAGRKRGWGWGKKGGWKKTTKTKKKKTDKTKTGTDTKWGWKDKWTDKKRDGKRGD 682

Qy        244 AAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATT 303
              :::::|     : : :|:: : :: :      |:||   |||:  |:| : ||  | ||
Db        683 GDRKDKGTGGDGDSKAKGDRAWAWRAKATRAAAADATAAATAAKATAWATKTTTKKTTTT 742

Qy        304 TCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATA 363
              | |:   | |  | |::  |  :|:::: :|::: ::: :::: :: : || : :::|:
Db        743 TTTKTTTTTTTTTTTKKTTTARKADRDDDDKAKRDGDKDGDWDDADKDWKAGRDRDDDTR 802

Qy        364 GGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGA 410
              :   : | :  |::: ::| :::|::| |  | :   | | | :|
Db        803 KWGWKDRAGKKKKAKKKDHKTAWWWADDTGTAWATWAKTATAKTGKA 849


RESULT 8
CL101205/c
LOCUS       CL101205               1072 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-38C6_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-38C6,
            genomic survey sequence.
ACCESSION   CL101205
VERSION     CL101205.1  GI:40594840
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1072)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 5
            High quality sequence stop: 744.
FEATURES             Location/Qualifiers
     source          1. .1072
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-38C6"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             9.9%;  Score 51;  DB 10;  Length 1072;
  Best Local Similarity   45.3%;  Pred. No. 0.23;
  Matches  186;  Conservative    0;  Mismatches  225;  Indels    0;  Gaps    0;

Qy         10 TATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAA 69
              || |  |  | ||| |||||||||| |||    || | |||    || |  ||      ||
Db        942 TAAAAATATAAATTTAAATTAAATTTTAAACAATAAATATCATATATATATAATATAA 883

Qy         70 TCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAAT 129
              |  | |  | |   ||   || |  ||     | |  |||| |||||       ||
Db        882 TATATATTATAAAATCACTTTATTATTTATTTAAACTATTTAAAAATTAAATATAATTTA 823

Qy        130 TTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGT 189
              | |||||   ||   | |   |  | |   | |||||| |  | |    |    ||
Db        822 TATAAATAATATAATATATTATTTTTATTATATAAATAATATATTTATATATATATATA 763

Qy        190 ATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTT 249
                  |||  |         |  |  | ||||||  ||     ||  | |  |
Db        762 TATATAAATTTATATATATATATATATATATATATATATATATACATATATATATATA 703

Qy        250 GAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGC 309
              |    ||| | ||||||  |      ||||| | ||  |    || |  |||| | |
Db        702 TATATATATATATACATATATATATATATATATATACATATATATATATATATATATA 643

Qy        310 GTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGAT 369
                 || | ||  | || | |   ||||||  ||   ||   |  | ||||||
Db        642 TACATATATATATATATATATATATATATATATACATATATATATATATATATATATA 583

Qy        370 CATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA 420
              ||||   ||     |   |||  ||  || |      || | ||||
Db        582 CATATATATATATATATATATATATAGATAGATAGATCTGCCCCAAGCA 532


RESULT 9
AW481292
LOCUS       AW481292               417 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  35294 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW481292
VERSION     AW481292.1  GI:7051398
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
```

```
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
            Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
            Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F.,
            Quackenbush,J. and Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 17  row: L  column: 23
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .417
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 1BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
ORIGIN

  Query Match             9.8%;  Score 50.6;  DB 1;  Length 417;
  Best Local Similarity   52.6%;  Pred. No. 0.26;
  Matches  110;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy        132 TAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTAT 191
              | |||| | | ||    | | | ||| |||| ||      || | | | |   ||
Db        161 TTAATTCATTTTCCTGACATTCACTGTAAAACATTCAGGGTCCTACCATTTCTGTTCAGG 220

Qy        192 GCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGA 251
              |||  | | | || |   |   ||   ||  | | | |    |||    |||| |
Db        221 AAATCTTGTAGTTTATTGTTATTTAACAGTATTAAGTGAATTTTTGTATGTTTCATTTTA 280

Qy        252 ACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGT 311
              || ||||||||  ||||   |    ||||||  |  || || || ||||||| ||
Db        281 ACTTTATTTGTGAATAGTGATTGTGGCATGTTTGGGGTATTATTTTAATATTTCATGATA 340

Qy        312 ATCATGTTTAATTTTAGAAAACGATATAT 340
              | | |||| | |||  ||||  || | |
Db        341 CTGAAGTTTTAATTTTTAAAAAAATTTCT 369


RESULT 10
CO884183
LOCUS       CO884183                599 bp    mRNA    linear   EST 01-SEP-2004
DEFINITION  BovGen_12508 normal cattle brain Bos taurus cDNA clone
            RZPDp1056M206Q 5', mRNA sequence.
ACCESSION   CO884183
VERSION     CO884183.1  GI:51814131
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Hennig,S., Janitz,M., Herwig,R. and Williams,J.
  TITLE     Generation, annotation, evolutionary analysis and database
            integration of 14969 cattle EST clusters
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Hennig S
            laboraty 123, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1612
            Fax: +49 30 8413 1380
            Email: hennig@molgen.mpg.de
            The library was characterised by oligonucleotide fingerprinting
            (ONFP) to reduce sequencing redundancy. According to the ONFP
            procedure, clones that display the same hybridisation matrix with a
            battery of 200 8mer oligonucleotides are grouped into clusters. One
            clone per ONFP cluster was selected for sequencing. cDNA clones and
            filters are distributed via Deutsches Ressourcenzentrum fuer
            Genomforschung GmbH (http://www.rzpd.de).
            PCR PRimers
            FORWARD: 5' CCCCAGGCTTTACACTTTATGCTTCCGGCTCG 3' (M13RSP) 5'-seq
            BACKWARD: 5' GCTATTACGCCAGCTGGCGAAAGGGGGATGTG 3' (M13FSP) 3'-seq
            Seq primer: 5'-CCGGTCCGGAATTCCCGGGT-3' (M13RSP).
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="RZPDp1056M206Q"
                     /sex="female"
                     /tissue_type="brain tissue"
                     /dev_stage="adult brain"
                     /clone_lib="normal cattle brain"
                     /note="Organ: brain; Vector: pSport1; Site_1: NotI;
                     Site_2: SalI; Random primed and directionally cloned in
                     pSport1 vector using NotI
                     (5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC (T)15-3' and SalI 5'-
                     TCGACCCACGCGTCCG-3' adapters (Gibco BRL)"
ORIGIN

  Query Match             9.8%;  Score 50.6;  DB 7;  Length 599;
  Best Local Similarity   52.6%;  Pred. No. 0.27;
  Matches  110;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy        132 TAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTAT 191
              | |||| | | ||    | | | ||| |||| ||      || | | | |   ||
Db         98 TTAATTCATTTTCCTGACATTCACTGTAAAACATTCAGGGTCCTACCATTTCTGTTCAGG 157

Qy        192 GCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGA 251
              |||  | | | || |   |   ||   ||  | | | |    |||    |||| |
Db        158 AAATCTTGTAGTTTATTGTTATTTAACAGTATTAAGTGAATTTTTGTATGTTTCATTTTA 217

Qy        252 ACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGT 311
              || ||||||||  ||||   |    ||||||  |  || || || ||||||| ||
Db        218 ACTTTATTTGTGAATAGTGATTGTGGCATGTTTGGGGTATTATTTTAATATTTCATGATA 277

Qy        312 ATCATGTTTAATTTTAGAAAACGATATAT 340
              | | |||| | |||  ||||  || | |
Db        278 CTGAAGTTTTAATTTTTAAAAAAATTTCT 306


RESULT 11
CB167163/c
LOCUS       CB167163                603 bp    mRNA    linear   EST 30-JAN-2003
DEFINITION  GWM602702549.R1 CSEQFXL25 muscle, rib eye Bos taurus cDNA, mRNA
            sequence.
ACCESSION   CB167163
VERSION     CB167163.1  GI:28153288
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Adelson,D.L. and Gill,C.A.
  TITLE     Bovine ESTs (Adelson and Gill)
```

```
  JOURNAL   Unpublished (2003)
COMMENT     Contact: David L. Adelson
            Animal Breeding and Genetics
            Texas A&M University
            Animal Science Dept., TAMU-2471, College Station, TX 77843-2471,
            USA
            Tel: 9798452616
            Fax: 9798456970
            Email: david.adelson@tamu.edu.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="muscle, rib eye"
                     /clone_lib="CSEQFXL25 muscle, rib eye"
                     /note="Organ: muscle, rib eye; Vector: pBluescript SK+;
                     Site_1: NotI; Site_2: EcoRI; sequence 5' of the insert
                     (5'-NNN. . .NNNinsert)
                     GCGAATTGGAGCTCCACCGCGGTGGCGGCCGCGGCTCGAG. Sequence 3' of
                     the inserts (AAGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAG.
                     non-normalized  library, sequenced 3' with M13R primer."
ORIGIN

  Query Match             9.8%;  Score 50.6;  DB 6;  Length 603;
  Best Local Similarity   52.6%;  Pred. No. 0.27;
  Matches  110;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy        132 TAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTAT 191
              | |||| | | ||    | | | ||| |||| ||      || | | | |   ||
Db        430 TTAATTCATTTTCCTGACATTCACTGTAAAACATTCAGGGTCCTACCATTTCTGTTCAGG 371

Qy        192 GCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGA 251
                |||  | | | || |   |   ||  ||  | | | |    |||    |||| |
Db        370 AAATCTTGTAGTTTATTGTTATTTAACAGTATTAAGTGAATTTTTGTATGTTTCATTTTA 311

Qy        252 ACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGT 311
              || |||||||  ||||    |     ||||||  |  || || || |||||||  ||
Db        310 ACTTTATTTGTGAATAGTGATTGTGGCATGTTTGGGGTATTATTTTAATATTTCATGATA 251

Qy        312 ATCATGTTTAATTTTAGAAAACGATATAT 340
                | | |||| | |||  ||||  || | |
Db        250 CTGAAGTTTTAATTTTTAAAAAAATTTCT 222


RESULT 12
AM027065/c
LOCUS       AM027065                659 bp    mRNA    linear   EST 30-JUN-2005
DEFINITION  AM027065 KN-252-lymph, Bos taurus Bos taurus cDNA clone C0007398c19
            3', mRNA sequence.
ACCESSION   AM027065
VERSION     AM027065.1  GI:68425954
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Anderson,S.I., Taylor,R., Talbot,R., Speed,D., Law,A.S.,
            Humphray,S., Hanotte,O., Mwakaya,J. and Archibald,A.L.
  TITLE     Development of cDNA and EST resources for studying host responses
            to trypanosome infection in cattle
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Anderson SI
            Genomics and Bioinformatics
            Roslin Institute
            Roslin, Midlothian, EH25 9PS, UNITED KINGDOM.
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="C0007398c19"
                     /tissue_type="lymph node"
                     /clone_lib="KN-252-lymph, Bos taurus"
                     /note="breed: N'dama"
ORIGIN

  Query Match             9.8%;  Score 50.6;  DB 1;  Length 659;
  Best Local Similarity   52.6%;  Pred. No. 0.27;
  Matches  110;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy        132 TAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTAT 191
              | |||| | | ||    | | | ||| |||| ||      || | | | |   ||
Db        404 TTAATTCATTTTCCTGACATTCACTGTAAAACATTCAGGGTCCTACCATTTCTGTTCAGG 345

Qy        192 GCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGA 251
                |||  | | | || |   |   ||  ||  | | | |    |||    |||| |
Db        344 AAATCTTGTAGTTTATTGTTATTTAACAGTATTAAGTGAATTTTTGTATGTTTCATTTTA 285

Qy        252 ACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGT 311
              || |||||||  ||||    |     ||||||  |  || || || |||||||  ||
Db        284 ACTTTATTTGTGAATAGTGATTGTGGCATGTTTGGGGTATTATTTTAATATTTCATGATA 225

Qy        312 ATCATGTTTAATTTTAGAAAACGATATAT 340
                | | |||| | |||  ||||  || | |
Db        224 CTGAAGTTTTAATTTTTAAAAAAATTTCT 196


RESULT 13
AM026437
LOCUS       AM026437                662 bp    mRNA    linear   EST 30-JUN-2005
DEFINITION  AM026437 KN-252-lymph, Bos taurus Bos taurus cDNA clone C0007398c19
            5', mRNA sequence.
ACCESSION   AM026437
VERSION     AM026437.1  GI:68425200
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 662)
  AUTHORS   Anderson,S.I., Taylor,R., Talbot,R., Speed,D., Law,A.S.,
            Humphray,S., Hanotte,O., Mwakaya,J. and Archibald,A.L.
  TITLE     Development of cDNA and EST resources for studying host responses
            to trypanosome infection in cattle
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Anderson SI
            Genomics and Bioinformatics
            Roslin Institute
            Roslin, Midlothian, EH25 9PS, UNITED KINGDOM.
FEATURES             Location/Qualifiers
     source          1. .662
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="C0007398c19"
                     /tissue_type="lymph node"
                     /clone_lib="KN-252-lymph, Bos taurus"
                     /note="breed: N'dama"
ORIGIN

  Query Match             9.8%;  Score 50.6;  DB 1;  Length 662;
  Best Local Similarity   52.6%;  Pred. No. 0.27;
  Matches  110;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy        132 TAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTAT 191
              | |||| | | ||    | | | ||| |||| ||      || | | | |   ||
Db        256 TTAATTCATTTTCCTGACATTCACTGTAAAACATTCAGGGTCCTACCATTTCTGTTCAGG 315

Qy        192 GCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGA 251
                |||  | | | || |   |   ||  ||  | | | |    |||    |||| |
Db        316 AAATCTTGTAGTTTATTGTTATTTAACAGTATTAAGTGAATTTTTGTATGTTTCATTTTA 375
```

```
Qy        252 ACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGT 311
              || |||||||| |||| |  |||||| | || || || ||||||| ||
Db        376 ACTTTATTTGTGAATAGTGATTGTGGCATGTTTGGGGTATTATTTTAATATTTCATGATA 435

Qy        312 ATCATGTTTAATTTTAGAAAACGATATAT 340
              | | |||| | |||  |||| || | |
Db        436 CTGAAGTTTTAATTTTTAAAAAAATTTCT 464


RESULT 14
CO891008
LOCUS       CO891008                663 bp    mRNA    linear   EST 01-SEP-2004
DEFINITION  BovGen_19333 normal cattle brain Bos taurus cDNA clone
            RZPDp1056P1339Q 5', mRNA sequence.
ACCESSION   CO891008
VERSION     CO891008.1  GI:51821301
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   Hennig,S., Janitz,M., Herwig,R. and Williams,J.
  TITLE     Generation, annotation, evolutionary analysis and database
            integration of 14969 cattle EST clusters
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Hennig S
            laboraty 123, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1612
            Fax: +49 30 8413 1380
            Email: hennig@molgen.mpg.de
            The library was characterised by oligonucleotide fingerprinting
            (ONFP) to reduce sequencing redundancy. According to the ONFP
            procedure, clones that display the same hybridisation matrix with a
            battery of 200 8mer oligonucleotides are grouped into clusters. One
            clone per ONFP cluster was selected for sequencing. cDNA clones and
            filters are distributed via Deutsches Ressourcenzentrum fuer
            Genomforschung GmbH (http://www.rzpd.de).
            PCR PRimers
            FORWARD: 5' CCCCAGGCTTTACACTTTATGCTTCCGGCTCG 3' (M13RSP) 5'-seq
            BACKWARD: 5' GCTATTACGCCAGCTGGCGAAAGGGGGATGTG 3' (M13FSP) 3'-seq
            Seq primer: 5'-CCGGTCCGGAATTCCCGGGT-3' (M13RSP).
FEATURES              Location/Qualifiers
     source           1. .663
                      /organism="Bos taurus"
                      /mol_type="mRNA"
                      /db_xref="taxon:9913"
                      /clone="RZPDp1056P1339Q"
                      /sex="female"
                      /tissue_type="brain tissue"
                      /dev_stage="adult brain"
                      /clone_lib="normal cattle brain"
                      /note="Organ: brain; Vector: pSport1; Site_1: NotI;
                      Site_2: SalI; Random primed and directionally cloned in
                      pSport1 vector using NotI
                      (5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC (T)15-3' and SalI 5'-
                      TCGACCCACGCGTCCG-3' adapters (Gibco BRL)"
ORIGIN

  Query Match             9.8%;  Score 50.6;  DB 7;  Length 663;
  Best Local Similarity   52.6%;  Pred. No. 0.27;
  Matches  110;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy        132 TAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTAT 191
              | |||| | | ||    | | | ||| |||| ||       || | | | |   ||
Db        210 TTAATTCATTTTCCTGACATTCACTGTAAAACATTCAGGGTCCTACCATTTCTGTTCAGG 269

Qy        192 GCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGA 251
              |||  | | | || |   |   ||   ||  | | | |    |||    |||| |
Db        270 AAATCTTGTAGTTTATTGTTATTTAACAGTATTAAGTGAATTTTTGTATGTTTCATTTTA 329

Qy        252 ACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGT 311
              || |||||||| |||| |  |||||| | || || || ||||||| ||
Db        330 ACTTTATTTGTGAATAGTGATTGTGGCATGTTTGGGGTATTATTTTAATATTTCATGATA 389

Qy        312 ATCATGTTTAATTTTAGAAAACGATATAT 340
              | | |||| | |||  |||| || | |
Db        390 CTGAAGTTTTAATTTTTAAAAAAATTTCT 418


RESULT 15
AM005285/c
LOCUS       AM005285                679 bp    mRNA    linear   EST 29-JUN-2005
DEFINITION  AM005285 KN-252-bone-marrow, Bos taurus Bos taurus cDNA clone
            C0005920112 3', mRNA sequence.
ACCESSION   AM005285
VERSION     AM005285.1  GI:68306476
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 679)
  AUTHORS   Anderson,S.I., Taylor,R., Talbot,R., Speed,D., Law,A.S.,
            Humphray,S., Hanotte,O., Mwakaya,J. and Archibald,A.L.
  TITLE     Development of cDNA and EST resources for studying host responses
            to trypanosome infection in cattle
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Anderson SI
            Genomics and Bioinformatics
            Roslin Institute
            Roslin, Midlothian, EH25 9PS, UNITED KINGDOM.
FEATURES              Location/Qualifiers
     source           1. .679
                      /organism="Bos taurus"
                      /mol_type="mRNA"
                      /db_xref="taxon:9913"
                      /clone="C0005920112"
                      /tissue_type="bone marrow"
                      /clone_lib="KN-252-bone-marrow, Bos taurus"
                      /note="breed: N'dama"
ORIGIN

  Query Match             9.8%;  Score 50.6;  DB 1;  Length 679;
  Best Local Similarity   52.6%;  Pred. No. 0.27;
  Matches  110;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy        132 TAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTAT 191
              | |||| | | ||    | | | ||| |||| ||       || | | | |   ||
Db        424 TTAATTCATTTTCCTGACATTCACTGTAAAACATTCAGGGTCCTACCATTTCTGTTCAGG 365

Qy        192 GCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGA 251
              |||  | | | || |   |   ||   ||  | | | |    |||    |||| |
Db        364 AAATCTTGTAGTTTATTGTTATTTAACAGTATTAAGTGAATTTTTGTATGTTTCATTTTA 305

Qy        252 ACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGT 311
              || |||||||| |||| |  |||||| | || || || ||||||| ||
Db        304 ACTTTATTTGTGAATAGTGATTGTGGCATGTTTGGGGTATTATTTTAATATTTCATGATA 245

Qy        312 ATCATGTTTAATTTTAGAAAACGATATAT 340
              | | |||| | |||  |||| || | |
Db        244 CTGAAGTTTTAATTTTTAAAAAAATTTTT 216


Search completed: May 12, 2006, 03:29:33
Job time : 1614.57 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 19:15:19 ; Search time 70.4361 Seconds
(without alignments)
12996.800 Million cell updates/sec

Title: US-09-423-093B-3_COPY_3361_3875
Perfect score: 515
Sequence: 1 atgtttgtctatagtttaag.....gctggtagtgaaatttgtga 515

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	43.4	8.4	1664976	3	US-08-916-421B-1	Sequence 1, Appli
	2	43.4	8.4	1664976	3	US-09-692-570-1	Sequence 1, Appli
	3	43	8.3	29717	3	US-09-949-016-16284	Sequence 16284, A
	4	43	8.3	60376	3	US-09-949-016-12423	Sequence 12423, A
c	5	42.6	8.3	640681	3	US-09-790-988-1	Sequence 1, Appli
c	6	42.2	8.2	57507	3	US-09-949-016-15019	Sequence 15019, A
	7	41.2	8.0	76269	3	US-09-949-016-14603	Sequence 14603, A
c	8	41	8.0	601	3	US-09-949-016-186365	Sequence 186365,
	9	41	8.0	91062	3	US-09-949-016-13019	Sequence 13019, A
	10	41	8.0	95255	3	US-09-949-016-17067	Sequence 17067, A
	11	40.8	7.9	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	12	40.6	7.9	251672	3	US-09-949-016-17296	Sequence 17296, A
c	13	40.6	7.9	251682	3	US-09-949-016-11973	Sequence 11973, A
	14	40.4	7.8	2916	3	US-09-248-796A-5428	Sequence 5428, Ap
	15	40.4	7.8	11165	3	US-09-830-807-34	Sequence 34, Appl
	16	40.4	7.8	32392	3	US-09-662-254B-27	Sequence 27, Appl
	17	40.2	7.8	483	3	US-09-710-279-2371	Sequence 2371, Ap
	18	40.2	7.8	601	3	US-09-949-016-137087	Sequence 137087,
	19	40.2	7.8	601	3	US-09-949-016-137088	Sequence 137088,
	20	40.2	7.8	3410	3	US-09-710-279-3505	Sequence 3505, Ap
	21	40.2	7.8	59948	3	US-09-949-016-15594	Sequence 15594, A
c	22	40.2	7.8	107800	3	US-09-949-016-13118	Sequence 13118, A
c	23	40.2	7.8	116425	3	US-09-949-016-11809	Sequence 11809, A
	24	40	7.8	1758	3	US-09-710-279-2123	Sequence 2123, Ap
	25	40	7.8	1806	3	US-09-134-001C-954	Sequence 954, App
	26	40	7.8	1830	3	US-09-662-254B-67	Sequence 67, Appl
c	27	40	7.8	2797	3	US-09-710-279-4369	Sequence 4369, Ap
	28	40	7.8	3100	3	US-09-710-279-4063	Sequence 4063, Ap
c	29	40	7.8	32392	3	US-09-662-254B-27	Sequence 27, Appl
	30	39.8	7.7	640681	3	US-09-790-988-1	Sequence 1, Appli
	31	39.6	7.7	3272	3	US-09-796-348-18	Sequence 18, Appl
	32	39.6	7.7	3303	3	US-09-796-348-19	Sequence 19, Appl
	33	39.6	7.7	3315	3	US-09-796-348-17	Sequence 17, Appl
c	34	39.6	7.7	21561	3	US-09-949-002-641	Sequence 641, App
c	35	39.6	7.7	21561	3	US-09-949-002-797	Sequence 797, App
	36	39.4	7.7	1167	3	US-09-248-796A-4678	Sequence 4678, Ap
c	37	39.4	7.7	50000	3	US-09-662-254B-25	Sequence 25, Appl
c	38	39.4	7.7	678533	3	US-09-949-016-14577	Sequence 14577, A
c	39	39.4	7.7	678533	3	US-09-949-016-14578	Sequence 14578, A
c	40	39	7.6	1050	3	US-09-662-254B-31	Sequence 31, Appl
	41	39	7.6	50000	3	US-09-662-254B-24	Sequence 24, Appl
	42	38.6	7.5	900	3	US-09-662-254B-57	Sequence 57, Appl
	43	38.6	7.5	4441	3	US-09-796-348-20	Sequence 20, Appl
	44	38.6	7.5	50000	3	US-09-662-254B-26	Sequence 26, Appl
	45	38.6	7.5	76281	3	US-09-949-016-12708	Sequence 12708, A

ALIGNMENTS

RESULT 1
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 8.4%; Score 43.4; DB 3; Length 1664976;
Best Local Similarity 49.3%; Pred. No. 0.57;
Matches 113; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```
Qy        287 GTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTG 346
              | ||||||  |   |||  ||   | |||   |||  |      ||  | |||  |||||
Db     466770 GAGATAATTGTGGAATTGGTTCATTTTCAATAATAAATGCTACTAAAAAAATAACAATTG 466829

Qy        347 GTAGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAA 406
              |||| || ||||  ||      ||||   |||   | | ||  ||       |    ||||
Db     466830 GTAGTAATGTTTTAATTTCAAGTCATGTTCATATTATTGATGGAGACCATGGATTTAAAA 466889

Qy        407 TAGAACCGCCAGCAAATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCT 466
                || |     |  ||  || |    |   |     |  |||||| |||| ||  ||
```

```
Db     466890 AAGGAGAATTAATAAGGAATCAGAAAATGGTCTCAGAGCCTATTGAAATTGGAGATGATG 466949

Qy        467 GCTGGATTGGAGATAATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 515
              |||||||||     |    |||  |    ||   | ||||| | ||
Db     466950 TTTGGATTGGAACAGGAGTTAAAATATTAAAAGGGGTTAAAATTGGGGA 466998


RESULT 2
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6797466
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275C1
;  CURRENT APPLICATION NUMBER: US/09/692,570
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  PRIOR APPLICATION NUMBER: US 08/916,421
;  PRIOR FILING DATE: 1997-08-22
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g

  Query Match             8.4%;  Score 43.4;  DB 3;  Length 1664976;
  Best Local Similarity   49.3%;  Pred. No. 0.57;
  Matches  113;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        287 GTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTG 346
              | ||||||  |   |||  ||   | |||    ||| |      ||  | |||  |||||
Db     466770 GAGATAATTGTGGAATTGGTTCATTTTCAATAATAAATGCTACTAAAAAAATAACAATTG 466829

Qy        347 GTAGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAA 406
              |||| || ||||  ||      ||||  |||   | | ||  ||      |   ||||
Db     466830 GTAGTAATGTTTTAATTTCAAGTCATGTTCATATTATTGATGGAGACCATGGATTTAAAA 466889

Qy        407 TAGAACCGCCAGCAAATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCT 466
              || |      |  ||  || |    |   |      |  |||||| |||| ||  ||
Db     466890 AAGGAGAATTAATAAGGAATCAGAAAATGGTCTCAGAGCCTATTGAAATTGGAGATGATG 466949

Qy        467 GCTGGATTGGAGATAATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 515
              |||||||||      |   |||  |   ||   | ||||| | ||
Db     466950 TTTGGATTGGAACAGGAGTTAAAATATTAAAAGGGGTTAAAATTGGGGA 466998


RESULT 3
US-09-949-016-16284
; Sequence 16284, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16284
;   LENGTH: 29717
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(29717)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16284

  Query Match             8.3%;  Score 43;  DB 3;  Length 29717;
  Best Local Similarity   44.9%;  Pred. No. 0.24;
  Matches  163;  Conservative    0;  Mismatches  200;  Indels    0;  Gaps    0;

Qy          1 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 60
              || | | | | ||  | |  | | ||  | |  |||   | ||   | ||  | |
Db      15474 ATATATATATTTATATATATTTATATATATATTTATATATATTTATATATATTTATATGT 15533

Qy         61 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 120
              |      ||  |    | | | |  | | | ||         |||| | ||  ||
Db      15534 ATATTTTTATATATTTATATATATTTATATATACATATTCATATATATTTTTATATATTT 15593

Qy        121 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 180
              | |  || | ||  | | | | |  || | |  | | | | | | |  | | |  |
Db      15594 ATATACATATTCATATATATTTTTATATATTTATATATACATATTTATATATATACATAT 15653

Qy        181 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 240
              |    ||          ||   |||     |    ||  |  |  ||||| |||| | ||
Db      15654 TTATATATATACATATTTATATATATACACATATTTATATATACATATATTTATATTTAT 15713

Qy        241 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 300
              | | ||  |   ||| | || |||  |      |||| | |   |  | || |  ||
Db      15714 ATATATTTATATTTATATATTTATATATTTATATTTATATATATTTATATTTATATATAT 15773

Qy        301 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT 360
              | | |        || ||||  |||| | |   ||||||   |   ||   | |||||
Db      15774 TTATATATTTATATATATTTATATTTATATATTTATATATATTTATATATTTATATTTAT 15833

Qy        361 ATA 363
              |||
Db      15834 ATA 15836
```

```
RESULT 4
US-09-949-016-12423
; Sequence 12423, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12423
;   LENGTH: 60376
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(60376)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12423

  Query Match             8.3%;  Score 43;  DB 3;  Length 60376;
  Best Local Similarity   44.9%;  Pred. No. 0.29;
  Matches  163;  Conservative    0;  Mismatches  200;  Indels    0;  Gaps    0;

Qy          1 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 60
              || | | | | ||  |  |  |  | ||  | |  |||    | ||   |   || |  |
Db      47513 ATATATATATTTATATATATTTATATATATATTTATATATATTTATATATATTTATATGT 47572

Qy         61 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 120
              |       ||  |     |  | | |   | | | ||         |||| | ||  ||
Db      47573 ATATTTTTATATATTTATATATATTTATATATACATATTCATATATATTTTTATATATTT 47632

Qy        121 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 180
              | |  || | ||  | | | | |  ||  | |  | | | | | |  | | | |  |
Db      47633 ATATACATATTCATATATATTTTTATATATTTATATATACATATTTATATATATACATAT 47692

Qy        181 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 240
              |    ||          ||    |||    |    ||  |  | ||||| ||||  | ||
Db      47693 TTATATATATACATATTTATATATATACACATATTTATATATACATATATTTATATTTAT 47752

Qy        241 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 300
              | | ||  |    ||| | || |||  |     ||||| | |   |  | || |  ||
Db      47753 ATATATTTATATTTATATATTTATATATTTATATTTATATATATTTATATTTATATATAT 47812

Qy        301 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT 360
              | | |         || ||||  |||| | |   ||||||   |   ||   |  |||||
Db      47813 TTATATATTTATATATATTTATATTTATATATTTATATATATTTATATATTTATATTTAT 47872

Qy        361 ATA 363
              |||
Db      47873 ATA 47875


RESULT 5
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             8.3%;  Score 42.6;  DB 3;  Length 640681;
  Best Local Similarity   48.2%;  Pred. No. 0.71;
  Matches  120;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        215 TTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGTGATC 274
              |||| || |    | | ||||  ||| |  |   | ||   | |  | |  ||     |
Db     465628 TTGATGCAGATATTGAATATGAAGATAATTATAATAAATGGTCTCAGGATTTATTACAAT 465569

Qy        275 GTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAACG 334
              |||   |   |  || ||  || |||  ||||  || |||| | ||  |  ||||||
Db     465568 CTATTAATAGTAAAGAAAAAATTTATAAATCTTCTGTGTCATATATAGATCAAGAAAATA 465509

Qy        335 ATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAAGTAT 394
                |  || ||   |  ||   ||||| |    |  ||      ||   ||  || | ||
Db     465508 CATTGATACTTTCGAAAAATCATTATACAAAAAAAAATCCTGCTGAAGGTATTATATTAA 465449

Qy        395 GCAGTCCGAAAATAGAACCGCCAGCAAATAAGCCATTAGGTGATATTGCTCCTATTAAAA 454
                | ||  |||||        |        ||||  |   | | ||  |  |   ||| | |
Db     465448 CAAATCAAAAAATTACTGGTCGTAATTCTAAGAAAGATGTTCATCATATTGAAATTGATA 465389

Qy        455 TAGGTAATT 463
              |  ||||||
Db     465388 TTAGTAATT 465380


RESULT 6
US-09-949-016-15019/c
; Sequence 15019, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15019
;   LENGTH: 57507
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15019

  Query Match             8.2%;  Score 42.2;  DB 3;  Length 57507;
  Best Local Similarity   44.8%;  Pred. No. 0.46;
  Matches  161;  Conservative    0;  Mismatches  198;  Indels    0;  Gaps    0;

Qy          5 TTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGC 64
              || |  |||  | || ||  | ||  ||   |||       ||   | | || |  || |
Db      28179 TTATACATATATAAATATATATATATAAAAATATATATAAATATATAAATATATATAAGT 28120

Qy         65 GGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGA 124
                | || || |   || | | |   | |   ||    |||| | | || ||   |    |
```

```
Db      28119 ATATATAAATATATAAATATATAAATATATATAATATATATAAATATATAAATATATATA 28060

Qy        125 AGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGG 184

Db      28059 ATATATATAAATATATAAATATATATAAATATATATGAATATATATAAATATAAATGAAT 28000

Qy        185 GTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAA 244

Db      27999 ATATATGAATATATATAAATATATATGAATATATATAAATATATAAGTATATATAAAT 27940

Qy        245 AATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTT 304

Db      27939 ATATATAAGTATATATGAATATATATAAATATATATTCATATATATAAATATATATTCAT 27880

Qy        305 CTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATA 363

Db      27879 ATATATAAATATATATTCATATATATAAATATATATGAATATATATATAAATATATATA 27821


RESULT 7
US-09-949-016-14603
; Sequence 14603, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14603
;   LENGTH: 76269
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-14603

  Query Match             8.0%;  Score 41.2;  DB 3;  Length 76269;
  Best Local Similarity   57.9%;  Pred. No. 0.91;
  Matches   73;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy         31 AATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAAGTTTCTTGTT 90

Db      56271 ATTCTTTCCTCAGCTATGTGAAGTCTACTAATGAGCCCATCAAAGGCATATTTTCTTTCT 56330

Qy         91 CTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAATGTCAAACCT 150

Db      56331 GTTTTTGATCTTTATCATTTTAAAATTATTTCCTAGAATTTTAATCTCTCTGCTTAAATT 56390

Qy        151 TACTCT 156

Db      56391 TCCTAT 56396


RESULT 8
US-09-949-016-186365/c
; Sequence 186365, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 186365
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-186365

  Query Match             8.0%;  Score 41;  DB 3;  Length 601;
  Best Local Similarity   44.0%;  Pred. No. 0.27;
  Matches  173;  Conservative    0;  Mismatches  220;  Indels    0;  Gaps    0;

Qy         65 GGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGA 124

Db        502 GTAATATAACAGAATTGTATATTATATACAAAATATAACAGAATTATATATTATATACAA 443

Qy        125 AGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGG 184

Db        442 AGTATATAGAATTATATATAATATATAATTATATATTATATACAAAATATAATATAGAAT 383

Qy        185 GTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAA 244

Db        382 TATATATTATATATGATTATATATTATATACAAAATATAACATAATATAGAATTATATAT 323

Qy        245 AATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTT 304

Db        322 AATATATGATTTTATATTATATACAAAATATAATATAGAATTATATTATATTTTAATTAT 263

Qy        305 CTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAG 364

Db        262 TAATATATATTTATTATATATTAATATATTATATTTAATAATTATTATATATTATATAT 203

Qy        365 GCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCAAATA 424

Db        202 TAATATATAATAATTATTAATATATAATATGTAATTATATATTATATACAAAATATAACA 143

Qy        425 AGCCATTAGGTGATATTGCTCCTATTAAAATAG 457

Db        142 ATTATATATTATATACAAAATATAATATAACAG 110


RESULT 9
US-09-949-016-13019
; Sequence 13019, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13019
;   LENGTH: 91062
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(91062)
;   OTHER INFORMATION: n = A,T,C or G
```

US-09-949-016-13019

Query Match 8.0%; Score 41; DB 3; Length 91062;
Best Local Similarity 46.1%; Pred. No. 1.1;
Matches 137; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

```
Qy         53 GTAAAGTTAGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTA 112
              ||||  ||    |||  |  | |   ||  ||      |    |||| ||  ||   |
Db       9592 GTAATTTTGCTGGGACTTGCAGAAATAACATTGAATCTGAAATTAGTGGGCATCCAATGG 9651

Qy        113 AAATGGTTGGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCT 172
              |     | |  |  |  | | ||     |    ||  |  | |  |    ||| | ||
Db       9652 AGTCACTAGAAATAATCTGTCAAAATCACTCAAATACAAATTTAGATTTAAAGACCAGGA 9711

Qy        173 CTTCCAAATGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATT 232
              | | || | |  | ||||    |   | |  | |   ||| | | |  |||| |  |
Db       9712 AATTCTAAAGAGTAACTATGGTAATATATTAACAATTTAGGAAAGGAACATTATACATAT 9771

Qy        233 ATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATA 292
              ||  | ||  |||  || |  | ||||| ||  ||       ||| | | ||  |   |
Db       9772 ATATTTATATAAATATTTATACATATTTATATTTATACATATATATTTATATATATATAA 9831

Qy        293 ATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTA 349
              || |  |||    ||     | || | ||| | || |||    ||||||  ||   ||
Db       9832 ATATATATAAACATTTATATTTATATATAAATATATAAACATTTATATTTATAAATA 9888
```

RESULT 10
US-09-949-016-17067
; Sequence 17067, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17067
; LENGTH: 95255
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(95255)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17067

Query Match 8.0%; Score 41; DB 3; Length 95255;
Best Local Similarity 44.0%; Pred. No. 1.1;
Matches 173; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

```
Qy         65 GGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGA 124
              | ||   || || |  || | || |       |   | |  ||| || | |   |   |
Db      42735 GTAATATAACAGAATTGTATATTATATACAAAATATAACAGAATTATATATTATATACAA 42794

Qy        125 AGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGG 184
              || || |  ||||  || | | |  | | | |  |  | |||  |  | |  | |
Db      42795 AGTATATAGAATTATATATAATATATAATTATATATTATATACAAAATATAATATAGAAT 42854

Qy        185 GTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAA 244
              |||      || | |||  |  |  |   ||  |      |  ||   | ||  |
Db      42855 TATATATTATATATGATTATATATTATATACAAAATATAACATAATATAGAATTATATAT 42914

Qy        245 AATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTT 304
              ||| |     ||  |  || |||    ||   ||||       |  ||   || |  | |
Db      42915 AATATATGATTTTATATTATATACAAAATATAATATAGAATTATATTATATTTTAATTAT 42974

Qy        305 CTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAG 364
                  | ||  || |||  | |  ||||||||||   ||  | |  |||||||
Db      42975 TAATATATATTTATTATATATTAATATATTATATATTAATAATTATTATATATTATATAT 43034

Qy        365 GCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCAAATA 424
                 |||   ||       |||| | |||| | |   | | || | |   |   || |
Db      43035 TAATATATAATAATTATTAATATATAATATGTAATTATATATTATATACAAAATATAACA 43094

Qy        425 AGCCATTAGGTGATATTGCTCCTATTAAAATAG 457
              |     ||    |||        || || || ||
Db      43095 ATTATATATTATATACAAAATATAATATAACAG 43127
```

RESULT 11
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 7.9%; Score 40.8; DB 3; Length 1141;
Best Local Similarity 9.8%; Pred. No. 0.37;
Matches 33; Conservative 156; Mismatches 146; Indels 1; Gaps 1;

```
Qy         32 ATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAAGTTTCTTGTTC 91
              :: :: :: ::|: ::::::: :: :: :::  |::: :::|:   :::::   :|:
Db       213 MBWWKNWSYDVTYYWWVWDDMCKRKVRRWVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMA 272

Qy         92 TGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAATGTCAAACCTT 151
              :: :|:      |: : :    : : :: :    :: :|  : |: :   |:|    :
Db       273 YRRWTMNNNNNNAKAMCKRAKYWGWNRABVNSTCTTWKSKTTKVRTSCWANNCRAGDANK 332

Qy        152 ACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGTGATAACTGCT 211
              ::: ::  ||: :::       :|:  |::: : :::::   :: :       :  : ::
Db       333 DHKWWKWSAAMGVYWNNNNNNNWTYKKARHBARWDWVWHSAWKKWHANAAHYSRKKWTBY 392

Qy        212 GGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGTG 271
              :: | :      |::  :: : :: : :::        | : :: :|: :::  : : :
Db       393 KRKTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYNNNNNGGRTYYGWTKNKKMWTYYKWKA 452

Qy        272 ATCGT-ATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAA 330
               |:: |:::| |:  | :: :: ::    :: :| | :::: | ::|: :::: : |:
Db       453 NNCKWRAWDHKTCTHNNTTWWKMKTYWNNCYWKSMTNGKSHRBAAAVYTWYMWWWRRYAH 512

Qy        331 AACGATATATTAATTGGTAGCAAAGTTTATATAGGC 366
              |    ::::::| |: :: :|  :   :| :: :::
Db       513 ANNNNWDYWWKACTWYKYBVCSKWWNNYAAWYTKSS 548
```

RESULT 12
US-09-949-016-17296/c
; Sequence 17296, Application US/09949016
; Patent No. 6812339

```
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17296
;   LENGTH: 251672
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(251672)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17296

  Query Match             7.9%;  Score 40.6;  DB 3;  Length 251672;
  Best Local Similarity   44.3%;  Pred. No. 1.8;
  Matches  166;  Conservative    0;  Mismatches  209;  Indels    0;  Gaps    0;

Qy          4 TTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGG 63
              | | | |||| ||| | ||   ||||| || |     ||| | |||  | ||||  ||
Db      62218 TATTTATATATTTTTATATATTAATTATATATATAATTATAAATATATATTAAATATATA 62159

Qy         64 CGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGG 123
                | ||  | |   | | | ||  | |   |  ||||  || || | |
Db      62158 TTTATATATATATTTATATATATTTATATATAGAATATATATATTTATATATAGAAATAT 62099

Qy        124 AAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGG 183
              |    | |  |    |  | |  |  | |  || | |||    |   |  ||
Db      62098 ATATTTATATATAGAATATATATATTTATATATAGAATATATATTTACATATAGAATATA 62039

Qy        184 GGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAA 243
              ||  |  |     |   |||  | ||  ||  |     ||||||  | |    ||  |
Db      62038 TATATTTTTATATATTTATATATATTCTATATATATAGAATATATATATTTATATATAGA 61979

Qy        244 AAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATT 303
              | ||||  |   ||| | || ||||   ||     || | |     | ||| |  ||  |
Db      61978 ATATTTATAGAATATATTTATATAGAATATATGGAATATATTCTATAAAATATATATTCT 61919

Qy        304 TCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATA 363
              |   |||  | | ||| |  |  | ||| | ||  |    | |  | |||||
Db      61918 ATATAAATATATTCTATAAAATATATATTCTATAAAATATATATATTCTATATAAATATA 61859

Qy        364 GGCGATCATAGCCAT 378
                | |||   ||
Db      61858 TATATTTATACTTAT 61844


RESULT 13
US-09-949-016-11973/c
; Sequence 11973, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11973
;   LENGTH: 251682
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(251682)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11973

  Query Match             7.9%;  Score 40.6;  DB 3;  Length 251682;
  Best Local Similarity   44.3%;  Pred. No. 1.8;
  Matches  166;  Conservative    0;  Mismatches  209;  Indels    0;  Gaps    0;

Qy          4 TTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGG 63
              | | | |||| ||| | ||   ||||| || |     ||| | |||  | ||||  ||
Db      62228 TATTTATATATTTTTATATATTAATTATATATATAATTATAAATATATATTAAATATATA 62169

Qy         64 CGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGG 123
                | ||  | |   | | | ||  | |   |  ||||  || || | |
Db      62168 TTTATATATATATTTATATATATTTATATATAGAATATATATATTTATATATAGAAATAT 62109

Qy        124 AAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGG 183
              |    | |  |    |  | |  |  | |  || | |||    |   |  ||
Db      62108 ATATTTATATATAGAATATATATATTTATATATAGAATATATATTTACATATAGAATATA 62049

Qy        184 GGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAA 243
              ||  |  |     |   |||  | ||  ||  |     ||||||  | |    ||  |
Db      62048 TATATTTTTATATATTTATATATATTCTATATATATAGAATATATATATTTATATATAGA 61989

Qy        244 AAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATT 303
              | ||||  |   ||| | || ||||   ||     || | |     | ||| |  ||  |
Db      61988 ATATTTATAGAATATATTTATATAGAATATATGGAATATATTCTATAAAATATATATTCT 61929

Qy        304 TCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATA 363
              |   |||  | | ||| |  |  | ||| | ||  |    | |  | |||||
Db      61928 ATATAAATATATTCTATAAAATATATATTCTATAAAATATATATATTCTATATAAATATA 61869

Qy        364 GGCGATCATAGCCAT 378
                | |||   ||
Db      61868 TATATTTATACTTAT 61854


RESULT 14
US-09-248-796A-5428
; Sequence 5428, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5428
;   LENGTH: 2916
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-09-248-796A-5428

  Query Match             7.8%;  Score 40.4;  DB 3;  Length 2916;
  Best Local Similarity   53.1%;  Pred. No. 0.6;
  Matches   86;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 60
              ||||| | ||     ||| | |||| || | |||| | || ||   ||| ||| |
Db       2384 ATGTTGATTTACGCAATAAATTGAAAACAAAGTATTATGAAATAATCCATGATGAAAAAT 2443

Qy         61 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 120
               ||  |  || || | |||   ||  | |||   |        || |||| | | |
Db       2444 TGGTTGTTATAAATACCAATTATTGGTTTTCAAATATAATTTTACGATTGTCAGAGTTAA 2503

Qy        121 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAA 162
                || |   ||| |||  |||| | | | |  | | |||||
Db       2504 TTCAATACAATTATATTTTATGTGATATCATGGTTTTCAAAA 2545


RESULT 15
US-09-830-807-34
; Sequence 34, Application US/09830807
; Patent No. 6846667
; GENERAL INFORMATION:
;  APPLICANT: Crooke,Helen R.
;  APPLICANT:  Clarke, Enda E.
;  APPLICANT:  Everest, Paul H.
;  APPLICANT:  Dougan, Gordon
;  APPLICANT:  Holden, David W.
;  APPLICANT:  Shea, Jacqueline E.
;  APPLICANT:  Feldman, Robert G.
;  TITLE OF INVENTION: VIRULENCE GENES AND PROTEINS, AND THEIR USE
;  FILE REFERENCE: GJE-65
;  CURRENT APPLICATION NUMBER: US/09/830,807
;  CURRENT FILING DATE:  2001-04-30
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
;   LENGTH: 11165
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3791)..(4834)
;   NAME/KEY: CDS
;   LOCATION: (10459)..(10776)
;   NAME/KEY: CDS
;   LOCATION: (10134)..(10427)
;   NAME/KEY: CDS
;   LOCATION: (9836)..(10081)
;   NAME/KEY: CDS
;   LOCATION: (7816)..(9480)
;   NAME/KEY: CDS
;   LOCATION: (4878)..(7802)
;   NAME/KEY: CDS
;   LOCATION: (3460)..(3702)
;   NAME/KEY: CDS
;   LOCATION: (3054)..(3407)
;   NAME/KEY: CDS
;   LOCATION: (2613)..(3041)
;   NAME/KEY: CDS
;   LOCATION: (2198)..(2530)
;   NAME/KEY: CDS
;   LOCATION: (1939)..(2196)
;   NAME/KEY: CDS
;   LOCATION: (1573)..(1893)
;   NAME/KEY: CDS
;   LOCATION: (1102)..(1485)
;   NAME/KEY: CDS
;   LOCATION: (2)..(1099)
US-09-830-807-34

  Query Match             7.8%;  Score 40.4;  DB 3;  Length 11165;
  Best Local Similarity   52.4%;  Pred. No. 0.87;
  Matches   89;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy          3 GTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAG 62
              | || | ||||   ||| ||||||    ||  | ||  ||||||||   |   | || |
Db       6495 GATTATGTATACAATAAAATTAAAGAACAAAATGATGCTATCATTAAAAATCTACGTAAT 6554

Qy         63 GCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGG 122
               | |      | ||  |||   ||| |   ||  |      | |   || |  |||   |||
Db       6555 TCTGTGGATATAAAGAAACCCTCTGGAGTTGGATTAACAAAAGAGTTATCTAATTTATGG 6614

Qy        123 GAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCT 172
               ||   || | |||| ||| | ||||   |||| |     |  | |  | ||
Db       6615 GATAGATATCAAATAGAAAGACAAAAAATACTGTTATCACTAAATGAGCT 6664


Search completed: May 12, 2006, 03:51:13
Job time : 79.4361 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:31:18 ; Search time 329.747 Seconds
(without alignments)
12915.142 Million cell updates/sec

Title: US-09-423-093B-3_COPY_3361_3875
Perfect score: 515
Sequence: 1 atgtttgtctatagtttaag.....gctggtagtgaaatttgtga 515

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	515	100.0	12441	3	US-09-423-093A-3	Sequence 3, Appli
	2	53.2	10.3	3673778	6	US-10-312-841-2	Sequence 2, Appli
c	3	49.6	9.6	408	9	US-10-275-323A-13	Sequence 13, Appl
	4	48.8	9.5	8711	7	US-10-221-714A-424	Sequence 424, App
	5	47.6	9.2	11422	6	US-10-311-455-191	Sequence 191, App
	6	47.6	9.2	11422	7	US-10-257-166-17	Sequence 17, Appl
	7	47.2	9.2	17967	6	US-10-311-455-987	Sequence 987, App
	8	46.8	9.1	6056	6	US-10-311-455-999	Sequence 999, App
	9	46.4	9.0	19634	8	US-10-473-126-155	Sequence 155, App
	10	46.4	9.0	19634	8	US-10-473-126-301	Sequence 301, App
	11	46.4	9.0	19659	6	US-10-311-455-740	Sequence 740, App
c	12	46	8.9	592	4	US-09-925-065A-491835	Sequence 491835,
c	13	46	8.9	592	4	US-09-925-065A-491836	Sequence 491836,
	14	46	8.9	15732	5	US-10-239-676-95	Sequence 95, Appl
	15	46	8.9	15732	6	US-10-240-453-107	Sequence 107, App
	16	46	8.9	17137	6	US-10-311-455-164	Sequence 164, App
	17	45	8.7	6087	6	US-10-311-455-391	Sequence 391, App
	18	44.2	8.6	13326	6	US-10-311-455-1685	Sequence 1685, Ap
	19	44	8.5	8711	7	US-10-221-714A-423	Sequence 423, App
	20	43.6	8.5	11047	5	US-10-239-676-188	Sequence 188, App
	21	43.6	8.5	11047	6	US-10-311-455-1958	Sequence 1958, Ap
	22	43.6	8.5	11047	6	US-10-240-453-276	Sequence 276, App
	23	43.4	8.4	6668	6	US-10-311-455-1192	Sequence 1192, Ap
	24	43.4	8.4	6668	7	US-10-221-714A-140	Sequence 140, App
	25	43.4	8.4	6775	7	US-10-433-793-190	Sequence 190, App
	26	43.4	8.4	46951	6	US-10-091-281-2	Sequence 2, Appli
c	27	43.4	8.4	3673778	6	US-10-312-841-1	Sequence 1, Appli
c	28	43.2	8.4	622	4	US-09-925-065A-780581	Sequence 780581,
	29	43.2	8.4	10329	6	US-10-311-455-2095	Sequence 2095, Ap
	30	43	8.3	8237	7	US-10-221-714A-527	Sequence 527, App
	31	43	8.3	14919	7	US-10-221-714A-228	Sequence 228, App
	32	43	8.3	3673778	6	US-10-312-841-1	Sequence 1, Appli
c	33	42.8	8.3	1183	4	US-09-925-065A-722804	Sequence 722804,
	34	42.8	8.3	6207	7	US-10-221-714A-504	Sequence 504, App
	35	42.6	8.3	6118	6	US-10-311-455-1003	Sequence 1003, Ap
	36	42.6	8.3	113515	6	US-10-311-455-2147	Sequence 2147, Ap
c	37	42.6	8.3	640681	3	US-09-790-988-1	Sequence 1, Appli
c	38	42.4	8.2	666	4	US-09-925-065A-61556	Sequence 61556, A
c	39	42.4	8.2	1350	9	US-10-450-763-8403	Sequence 8403, Ap
	40	42.4	8.2	6641	6	US-10-311-455-288	Sequence 288, App
	41	42.4	8.2	6641	6	US-10-240-452-36	Sequence 36, Appl
	42	42.4	8.2	19082	6	US-10-311-455-600	Sequence 600, App
	43	42	8.2	9106	6	US-10-240-452-69	Sequence 69, Appl
	44	41.8	8.1	40324	7	US-10-433-793-180	Sequence 180, App
	45	41.6	8.1	5995	7	US-10-221-714A-446	Sequence 446, App

ALIGNMENTS

RESULT 1
US-09-423-093A-3
; Sequence 3, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12441
; TYPE: DNA
; ORGANISM: Salmonella enterica
US-09-423-093A-3

Query Match 100.0%; Score 515; DB 3; Length 12441;
Best Local Similarity 100.0%; Pred. No. 2.8e-108;
Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 ATGTTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTT 3420

Qy         61 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 AGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTT 3480

Qy        121 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GGGAAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAA 3540

Qy        181 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 TGGGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGAT 3600

Qy        241 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 GAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCAT 3660
```

```
Qy        301 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 ATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTAT 3720

Qy        361 ATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 ATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCA 3780

Qy        421 AATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 AATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGCTGGATTGGAGAT 3840

Qy        481 AATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 515
              |||||||||||||||||||||||||||||||||||
Db       3841 AATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 3875


RESULT 2
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (379615)
US-10-312-841-2

  Query Match             10.3%;  Score 53.2;  DB 6;  Length 3673778;
  Best Local Similarity   49.5%;  Pred. No. 1.1;
  Matches  191;  Conservative    0;  Mismatches  193;  Indels    2;  Gaps    2;

Qy         12 TAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATC 71
              ||   |||  | |   | |  | |||| | |  ||||||   || |  |  |  | | |
Db      45704 TATAGTAAATTAAGGTTAATTTATTATAAAAGAATTATTTTTTATAAATTTGATGTAGTT 45763

Qy         72 AAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTT 131
               ||     |   | || |   | |||  | |     ||  |||  |  | ||     || |
Db      45764 TAATTGTATTATGTTTTTAAAGTTTATAGTAGTGTATAGTAAGGTTTTAGGTTTTTATAT 45823

Qy        132 TAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTAT 191
              | | ||     | |    ||| | |   | | |||    | ||  ||   |  | | |
Db      45824 TTATTTATTATTTATTTATTAATTTATTTAGAGTAATTTTTATTTAAATATGTTTTATTT 45883

Qy        192 GCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGA 251
                |||| ||  |   |   ||  ||     | | ||| |||     | ||      | |||
Db      45884 AATGGTTAGTATTTTGTATAGGTGTGTTATTTTTTATTATTTATATTATATTTTAATTGT 45943

Qy        252 ACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATT-TCTTGCG 310
              |  || ||| |   |||||      ||||| | ||| || | ||  |  | || | ||
Db      45944 ATTTTTTTTATGTTTAGGTATGTTTATATATATAAATGGTTATTATTGTGTTATGTTATT 46003

Qy        311 TATCATGTTTAATTTTAG-AAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGAT 369
              ||| |||||| | |  ||| || ||| |||||   || || |  |     |||||  ||
Db      46004 TATTATGTTTTAGTATAGTAATACGTTATATAGGTTTGTTGTTTAAGAATAATAGGTTAT 46063

Qy        370 CATAGCCATGGCAGTTATAAAGTATG 395
              | |||  |    || | |  |||| |
Db      46064 CGTAGATAGTTTAGGTGTGTAGTAGG 46089


RESULT 3
US-10-275-323A-13/c
; Sequence 13, Application US/10275323A
; Publication No. US20050079491A1
; GENERAL INFORMATION:
;  APPLICANT: DONNE-GOUSSE, CAROLE
;  APPLICANT:  LAUDET, VINCENT
;  APPLICANT:  HANNI, CATHERINE
;  TITLE OF INVENTION: METHOD OF DETECTING AND IDENTIFYING THE PRESENCE OF BIOLOGICAL
;  TITLE OF INVENTION:  MATERIALS DERIVED FROM BIRDS, AND OLIGONUCLEOTIDES FOR CARRYING
;  TITLE OF INVENTION:  IT OUT
;  FILE REFERENCE: 0508-1052
;  CURRENT APPLICATION NUMBER: US/10/275,323A
;  CURRENT FILING DATE:  2002-11-08
;  PRIOR APPLICATION NUMBER: PCT/FR01/01279
;  PRIOR FILING DATE: 2001-04-26
;  PRIOR APPLICATION NUMBER: FR 00/05850
;  PRIOR FILING DATE: 2000-05-09
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
;   LENGTH: 408
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: DNA fragment amplified
;   OTHER INFORMATION: from the genome of birds
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (43)..(43)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (57)..(57)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (71)..(71)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (134)..(134)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (212)..(212)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (214)..(214)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (217)..(217)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (226)..(227)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (236)..(236)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (245)..(245)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (253)..(253)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
```

```
;   LOCATION: (269)..(269)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (324)..(324)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (330)..(330)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (352)..(352)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (355)..(355)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (357)..(357)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (362)..(362)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (364)..(364)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (406)..(406)
;   OTHER INFORMATION: a, c, g, or t
US-10-275-323A-13

  Query Match             9.6%;  Score 49.6;  DB 9;  Length 408;
  Best Local Similarity   9.6%;  Pred. No. 0.2;
  Matches   28;  Conservative  168;  Mismatches   97;  Indels    0;  Gaps    0;

Qy        223 GTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGT 282
              :::::: :::  :  |:::::: :: : :   :| :    :  ::::| :| ::::::::
Db        391 RYRDDDBRDKYVRYTRRYDRDHYRDRBNVNDDKTNKNRANRDBDRRDTWRTRRKDDWYDY 332

Qy        283 TTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTA 342
              : :::|: :: ::::: |: | || | || :|     |||:::|:::: ::::::| :::
Db        331 YNRDKTDNDWBHDBBBYTRGTRCTGGWGTTYCTMTAATAADKWTRDDHBRHDHHRTKHDW 272

Qy        343 ATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCG 402
              :: ::::: ::  :: :  :: :::::  :::: :: :  ::::::  : : :|   ::
Db        271 WYNKKWHDGWHKHKHVRVNDDWDBVWNRDDWRRSRNRWTWDDDRNNDGVRRTWDNRDNDN 212

Qy        403 AAAATAGAACCGCCAGCAAATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAAT 462
              :::::: ::   :   : :: : ::  :::| :: |::: : :  ::::::|:::| ::|
Db        211 DDDDWWTDVDDKWDKDDDDKDBDDRDWWKAYRHWABRWRDWKKDWWDHWDAWYDGHKRAA 152

Qy        463 TGCTGCTGGATTGGAGATAATGCAGTAATTCTGGCTGGTAGTGAAATTTGTGA 515
              ::::||| |: :     :::  ::::   : :: : :::::: :::: :::
Db        151 RWHYDYTGGTTKHRTTYNKDDDKYSYRDRRRWWRYRSWDRRWRDKWWDWTKKW 99


RESULT 4
US-10-221-714A-424
; Sequence 424, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 424
;   LENGTH: 8711
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-424

  Query Match             9.5%;  Score 48.8;  DB 7;  Length 8711;
  Best Local Similarity   48.5%;  Pred. No. 1;
  Matches  165;  Conservative    0;  Mismatches  172;  Indels    3;  Gaps    1;

Qy         35 TTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAAGTTTCTTGTTCTGC 94
              || | ||   |  |||  ||     | ||  ||||  | |||  | | |   ||||
Db       2643 TTGTAATTATAAAATTTGGTTTTTGTTGGGAGAAAATAGAAGTGAGGATAAATGTTTATA 2702

Qy         95 TTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAATGTCAAACCTTACT 154
              || |   ||   |||||||  |  |  |     ||  ||  ||    |  |||  |||
Db       2703 TTGGATTATTAATTTTAAATTATTAAGATTTGATAATATTTTTGTAGGGAAAGATTATA 2762

Qy        155 CTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGTGATAACTGCTGGA 214
              ||  | | |||      | ||  || |  || |  |||  | ||| |   ||    || |
Db       2763 TTGTTATATATAGGTTATATTTTAATTTAGGGATGATGAAGTTTGATATGAATATATGTA 2822

Qy        215 TTGAAGCTGTATATAATTATGGTGATGAAAAATTTGA---ACCTTATTTGTACATAGGTG 271
              |  ||  |    ||   | |||| || | ||||||  |    |  |  |||||  |  | ||
Db       2823 TATAAATTTGTAATTTGTGTGGTAATTATAAATTTTAGTGATTTGGTTTGTTTAATGATG 2882

Qy        272 ATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAA 331
              |  |||    | |     |||     | |    ||   ||| | |||||  |   ||||
Db       2883 AATGTAGTGTTATGTTAAATATATAATAAACGATTTTTTTATTAAGTTTATTAAAAGAAT 2942

Qy        332 ACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCA 371
                 | | ||  |||| |   |  ||| | | | || | |
Db       2943 TTTTTTTTTTTTTTGGGATGGAGTTTTGTTTTGTCGTTTA 2982


RESULT 5
US-10-311-455-191
; Sequence 191, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 191
;   LENGTH: 11422
```

```
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-191

  Query Match             9.2%;  Score 47.6;  DB 6;  Length 11422;
  Best Local Similarity   48.0%;  Pred. No. 2.2;
  Matches  169;  Conservative    0;  Mismatches  179;  Indels    4;  Gaps    1;

Qy         24 AAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAAGTT 83
              || | || |  |||  || |  || || | ||    | | || ||  || |    |  |
Db       3873 AATAATATAGATTAAAATTTTGTTTTTTAATAGTTATTAGTGGTAAGAAATATAGTAAGT 3932

Qy         84 TCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAATGT 143
              | ||| | ||   |   ||  |||| |  || |  ||  |    | | ||| | | |
Db       3933 TTTTGATTTGTAGAAATTTTAGTTTTTTTATTTGATGAGGTGGTTAATAATATTAATTTT 3992

Qy        144 CAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGT-- 201
              | |  |         || | | ||  |   |  |  |||| |   | |||  |     |||||
Db       3993 TATATTTGTTGAAATAATATAAAATATTTAATATAAATTGTTAATTATAATGATTGGTAT 4052

Qy        202 --GATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATT 259
                | ||          ||  | | |   |||| || |      |||||| ||   |  |||||
Db       4053 AAGGTAGAATTATAATAAATGTTAATTATATTTGTTTGTTTGAAAATTTAATATTTTATT 4112

Qy        260 TGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTT 319
              |  ||    |  || ||| | ||| |  |||| |   | |||  |      | ||
Db       4113 AAAAGGTAAAAAAGTGTTTATTTGTAAATATTAATATATGTTTTTTGTTTAGTTTATAGA 4172

Qy        320 TAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCA 371
              | |||    | | || || | ||    | |      | | | |||||  ||| |
Db       4173 AAGTTTATAAGTAAGAGATTTAAAAGTGAAAGTTATTATTTATAGAAGATTA 4224


RESULT 6
US-10-257-166-17
; Sequence 17, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
;  TITLE OF INVENTION:  Genes Implicated in Pharmacogenomics
;  FILE REFERENCE: 5013.1011
;  CURRENT APPLICATION NUMBER: US/10/257,166
;  CURRENT FILING DATE:  2002-10-07
;  PRIOR APPLICATION NUMBER: PCT/EP01/07470
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-06-29
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 17
;   LENGTH: 11422
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-17

  Query Match             9.2%;  Score 47.6;  DB 7;  Length 11422;
  Best Local Similarity   48.0%;  Pred. No. 2.2;
  Matches  169;  Conservative    0;  Mismatches  179;  Indels    4;  Gaps    1;

Qy         24 AAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAAGTT 83
              || | || |  |||  || |  || || | ||    | | || ||  || |    |  |
Db       3873 AATAATATAGATTAAAATTTTGTTTTTTAATAGTTATTAGTGGTAAGAAATATAGTAAGT 3932

Qy         84 TCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAATGT 143
              | ||| | ||   |   ||  |||| |  || |  ||  |    | | ||| | | |
Db       3933 TTTTGATTTGTAGAAATTTTAGTTTTTTTATTTGATGAGGTGGTTAATAATATTAATTTT 3992

Qy        144 CAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGT-- 201
              | |  |         || | | ||  |   |  |  |||| |   | |||  |     |||||
Db       3993 TATATTTGTTGAAATAATATAAAATATTTAATATAAATTGTTAATTATAATGATTGGTAT 4052

Qy        202 --GATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATT 259
                | ||          ||  | | |   |||| || |      |||||| ||   |  |||||
Db       4053 AAGGTAGAATTATAATAAATGTTAATTATATTTGTTTGTTTGAAAATTTAATATTTTATT 4112

Qy        260 TGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTT 319
              |  ||    |  || ||| | ||| |  |||| |   | |||  |      | ||
Db       4113 AAAAGGTAAAAAAGTGTTTATTTGTAAATATTAATATATGTTTTTTGTTTAGTTTATAGA 4172

Qy        320 TAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCA 371
              | |||    | | || || | ||    | |      | | | |||||  ||| |
Db       4173 AAGTTTATAAGTAAGAGATTTAAAAGTGAAAGTTATTATTTATAGAAGATTA 4224
```

```
Db       3933 TTTTGATTTGTAGAAATTTTAGTTTTTTTATTTGATGAGGTGGTTAATAATATTAATTTT 3992

Qy        144 CAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGT-- 201
              | |  |         || | | ||  |   |  |  |||| |   | |||  |     |||||
Db       3993 TATATTTGTTGAAATAATATAAAATATTTAATATAAATTGTTAATTATAATGATTGGTAT 4052

Qy        202 --GATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATT 259
                | ||          ||  | | |   |||| || |      |||||| ||   |  |||||
Db       4053 AAGGTAGAATTATAATAAATGTTAATTATATTTGTTTGTTTGAAAATTTAATATTTTATT 4112

Qy        260 TGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTT 319
              |  ||    |  || ||| | ||| |  |||| |   | |||  |      | ||
Db       4113 AAAAGGTAAAAAAGTGTTTATTTGTAAATATTAATATATGTTTTTTGTTTAGTTTATAGA 4172

Qy        320 TAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCA 371
              | |||    | | || || | ||    | |      | | | |||||  ||| |
Db       4173 AAGTTTATAAGTAAGAGATTTAAAAGTGAAAGTTATTATTTATAGAAGATTA 4224


RESULT 7
US-10-311-455-987
; Sequence 987, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 987
;   LENGTH: 17967
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-987

  Query Match             9.2%;  Score 47.2;  DB 6;  Length 17967;
  Best Local Similarity   49.6%;  Pred. No. 3.2;
  Matches  121;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy         83 TTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAATG 142
              || ||||| |  ||    |  | |    || |||| |||   | |        | | | |
Db      13531 TTTTTGTTGTTTTTTAAGTGTTTTTAATTTAAATAGTTAATATGTTAGAAGAGTATATTT 13590

Qy        143 TCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGTG 202
              | | |  ||  |    | || | || |  |  |  ||||  |||| ||  |  ||  ||
Db      13591 TAATATTTTTATTAATATAATAAAAAATTTAATTATAATGTAGTAGAAATAAGATTTAATG 13650

Qy        203 ATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGT 262
                |  | |  | || || |||||  |||||   |      |||  | ||    ||| | |
Db      13651 TAAGTAGTTTTAATGTAGTTGTATTAAATTAAATTAGAAGTAAAATAGATGAATATATAT 13710

Qy        263 ACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAA 322
              | |||  | |   | |||  |   |  |||||  ||  |  ||   ||  |||||
Db      13711 ATATAATTTAAGTTTTATTTAGGATAGTATATGTTTATTATGTTTATAAATTGTGTTTTT 13770

Qy        323 TTTT 326
                |||
Db      13771 GTTT 13774


RESULT 8
```

```
US-10-311-455-999
; Sequence 999, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 999
;   LENGTH: 6056
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-999

  Query Match             9.1%;  Score 46.8;  DB 6;  Length 6056;
  Best Local Similarity   47.5%;  Pred. No. 2.6;
  Matches  172;  Conservative    0;  Mismatches  187;  Indels    3;  Gaps    1;

Qy          6 TGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCG 65
              | | |  |  | || |||  | ||   | |||  |||     ||| |  |||  |||
Db       3749 TTTTTCGAAATGAAAATTTTATTTTGTTTTTAAAATAAATAAATTAATAAAAAGTAGTGT 3808

Qy         66 GAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAA 125
                |    ||       |||| ||| | |    |      | || ||||| | |  ||  | |
Db       3809 TTATGTAAGTATTTGTTTTTTGATTTTAGGAATTTTTTGTTTTTATATTAATTTTGTA 3868

Qy        126 GAATTTTAAATTGAATGTCAAACCTTACTCTGCAAA---AAATAACACCTCTTCCAAATG 182
                |  ||  | ||  | |||  ||| | ||  ||      |   |        | ||
Db       3869 TTGGTAAAATATAAAAATAAAATGTTATTTTGTTAATTTTTTTTTTTAGTAAAGTAATATA 3928

Qy        183 GGGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGA 242
              |  |||||  |  |  |  | || || |  |   |  ||| | || | ||  |  | |
Db       3929 GTTTAGTATATGTTTAAGAAAAAAATGGTTAAGAAATTTTGTTTTTATTAATTATTTTTA 3988

Qy        243 AAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATAT 302
              | | | |||   |||| | |   |   | | |||     ||||  ||| || ||   ||
Db       3989 AGATTGTGATATTTATATTTTGTTTTTTAAACGTTATATTTTATAAGATTATTTTTTAAT 4048

Qy        303 TTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATAT 362
              | |||  ||| |  ||||| | ||   |    || |||   |  ||   |  ||    |
Db       4049 ATGTTGAATATAAGTTTTAAATATATGTATTTTTAAATTGTATTTTAAATATTTTAGGTT 4108

Qy        363 AG 364
              ||
Db       4109 AG 4110


RESULT 9
US-10-473-126-155
; Sequence 155, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Methods and nucleic acids for the analysis of  hematopoietic cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/473,126
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 155
;   LENGTH: 19634
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-155

  Query Match             9.0%;  Score 46.4;  DB 8;  Length 19634;
  Best Local Similarity   48.4%;  Pred. No. 5.1;
  Matches  187;  Conservative    0;  Mismatches  196;  Indels    3;  Gaps    2;

Qy          4 TTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGG 63
              ||||| | |||||| || ||| | |||  |    | |||| |||| | || || |  |
Db      11021 TTTGTTTTTAGTTTGAGTTTATAGTTATGTAAATTAATATTATTAGTTAGAAAGGGAAAT 11080

Qy         64 CGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGG 123
               || ||  |   |  |   || ||       |||   ||  |        ||    |||
Db      11081 AGGTAAAAATGTGTTATTGTTTTTTGGAAATTTATTAGAATTTTAAGGTAAGGAATTGAA 11140

Qy        124 AAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGG 183
              |       | | || ||  |  |     | | |   || ||||  |    ||   ||| |
Db      11141 ATTTTAGGTTATTTTAAGATTGATTAGGAATATTATAATAATATTAAAGATTGAGAATTG 11200

Qy        184 GGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAA 243
              | | ||  |   |  ||  || | |       |  ||    | | ||||||| | ||  ||
Db      11201 GATTGTGGGTTTGGGGGAAAT-ATTATATTTTGAAATGATTTTGTAATTATTGAGAATAA 11259

Qy        244 AAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATA-- 301
              | ||   |  |||    |  |||      |    ||||||  | ||||||  ||
Db      11260 ATATGGAATTTTTAGAATTTAATATAGATGGTTTATATTTTAAGATAAAATGTTGTTAAT 11319

Qy        302 TTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATA 361
              ||| ||   | |    ||| ||||||  |||   |  ||  |||| ||| ||   | |||
Db      11320 TTTGTTTTATTTATATTTTTATTTTAATAAATATTTGATAGATTGATAGGAATTATAATA 11379

Qy        362 TAGGCGATCATAGCCATGGCAGTTAT 387
              |    | |  |||  |    | || |
Db      11380 TGTTAGGTTGTAGTTAAATGAATTTT 11405


RESULT 10
US-10-473-126-301
; Sequence 301, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Methods and nucleic acids for the analysis of  hematopoietic cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/473,126
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 301
;   LENGTH: 19634
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-301

  Query Match             9.0%;  Score 46.4;  DB 8;  Length 19634;
  Best Local Similarity   48.4%;  Pred. No. 5.1;
  Matches  187;  Conservative    0;  Mismatches  196;  Indels    3;  Gaps    2;

Qy          4 TTTGTCTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGG 63
              ||||| | |||||| || ||| | |||  |    | |||| |||| | || || |  |
Db      11021 TTTGTTTTTAGTTTGAGTTTATAGTTATGTAAATTAATATTATTAGTTAGAAAGGGAAAT 11080

Qy         64 CGGAAATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGG 123
               || ||  |   |  |   || ||       |||   ||  |        ||    |||
Db      11081 AGGTAAAAATGTGTTATTGTTTTTTGGAAATTTATTAGAATTTTAAGGTAAGGAATTGAA 11140

Qy        124 AAGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGG 183
```

```
Db      11141 ATTTTAGGTTATTTTAAGATTGATTAGGAATATTATAATAATATTAAAGATTGAGAATTG 11200

Qy        184 GGTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAA 243

Db      11201 GATTGTGGGTTTGGGGGAAAT-ATTATATTTTGAAATGATTTTGTAATTATTGAGAATAA 11259

Qy        244 AAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATA-- 301

Db      11260 ATATGGAATTTTTAGAATTTAATATAGATGGTTTATATTTTAAGATAAAATGTTGTTAAT 11319

Qy        302 TTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATA 361

Db      11320 TTTGTTTTATTTATATTTTTATTTTAATAAATATTTGATAGATTGATAGGAATTATAATA 11379

Qy        362 TAGGCGATCATAGCCATGGCAGTTAT 387

Db      11380 TGTTAGGTTGTAGTTAAATGAATTTT 11405


RESULT 11
US-10-311-455-740
; Sequence 740, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 740
;   LENGTH: 19659
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-740

  Query Match             9.0%;  Score 46.4;  DB 6;  Length 19659;
  Best Local Similarity   46.8%;  Pred. No. 5.2;
  Matches  178;  Conservative    0;  Mismatches  201;  Indels    1;  Gaps    1;

Qy         20 GATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAA 79

Db       3238 GATGAGAGAAAAGTGGGAAATTAGATGGAATGAATAAGGTTTAAAGAGGTTTTGAAATTA 3297

Qy         80 AGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGA 139

Db       3298 TTTTTGTTGGAATATTTAGTTTTGGGTATAATAGGAAGTAAGGGAGGAAGTTTAAGAGGG 3357

Qy        140 ATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTG 199

Db       3358 ATATTGAGAAAGTTTAGAATTAAAATATAAAAATAATTTGTAGAGGAAGTTGTTGTATTA 3417

Qy        200 GTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATT 259

Db       3418 TGATTTTTTAGGTGAGTTATTTAAAATGTATTTTTGATAAAATAATTTTTATAAATTTATT 3477

Qy        260 TGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTT 319

Db       3478 TGTTTAAAGAGAAGATATTTAGTTTATAAGATAAAGTTAATGTATGTTAATTATTGT-TT 3536

Qy        320 TAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATG 379

Db       3537 TAATTTTTAAATATAGTAAATAAATATTAAAGTTAATTTGTAATGTGATTATTATTAATT 3596

Qy        380 GCAGTTATAAAGTATGCAGT 399

Db       3597 GATTTTTTATTTTATGGAAT 3616


RESULT 12
US-09-925-065A-491835/c
; Sequence 491835, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 491835
;   LENGTH: 592
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-491835

  Query Match             8.9%;  Score 46;  DB 4;  Length 592;
  Best Local Similarity   51.5%;  Pred. No. 1.6;
  Matches  106;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        198 TGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTA 257

Db        534 TGATAATGGGTGTGTGTGTGTGTGTGTGTATATATATATATATATATATATGATAATGTG 475

Qy        258 TTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATG 317

Db        474 TATATATATATATGATAATGTATATATATATGATAATGTGTATATATATATGATAATGTG 415

Qy        318 TTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCA 377

Db        414 TGTATATATATGATAATTTGCATATATATATGATAATGTGTATATATATATGTATATATACA 355

Qy        378 TGGCAGTTATAAAGTATGCAGTCCGA 403

Db        354 CATTAAAGTCACAGTATGCTCTGAGA 329


RESULT 13
US-09-925-065A-491836/c
; Sequence 491836, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
```

```
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 491836
;   LENGTH: 592
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-491836

  Query Match             8.9%;  Score 46;  DB 4;  Length 592;
  Best Local Similarity   51.5%;  Pred. No. 1.6;
  Matches  106;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        198 TGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTA 257
              || | ||   ||   |  ||    ||| ||||  |||    ||  | |  |   |  |
Db        534 TGATAATGGGTGTGTGTGTGTGTGTGTATATATATATATATATATATATGATAATGTG 475

Qy        258 TTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATG 317
              | | || |||  ||||  | ||| | ||  |||||||||  |||| | |    ||   ||
Db        474 TATATATATATATGATAATGTATATATATATGATAATGTGTATATATATATGATAATGTG 415

Qy        318 TTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCA 377
              | ||  | ||  | |   |  ||  ||   |   || || ||||||   |   |||  ||
Db        414 TGTATATATATGATAATTTGCATATATATATGATAATGTGTATATATATGTATATATACA 355

Qy        378 TGGCAGTTATAAAGTATGCAGTCCGA 403
                  |     | |||||||  |  ||
Db        354 CATTAAAGTCACAGTATGCTCTGAGA 329


RESULT 14
US-10-239-676-95
; Sequence 95, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 95
;   LENGTH: 15732
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-95

  Query Match             8.9%;  Score 46;  DB 5;  Length 15732;
  Best Local Similarity   47.7%;  Pred. No. 5.8;
  Matches  168;  Conservative    0;  Mismatches  180;  Indels    4;  Gaps    1;

Qy         22 TTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAAG 81
              ||   ||||  |  ||| || | |    |  ||||   |||  ||||
Db      13292 TTTTGATTACGTTATATTATTTAAGGGAGAAAGTTAAATAGGGATAAAGAAAAATATTTA 13351

Qy         82 TTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAAT 141
              || ||      |   | || || ||| |   ||||| ||||   ||||   |  |||
Db      13352 GAAGTTTTTTATTATTTTGAATGTGGTTTAATTTTGGTTTGGAATTTTTTTTGTTGTAAT 13411

Qy        142 GTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGT 201
                          || ||| | ||        ||  |||| |  |  |||  ||  |  |
Db      13412 TAATTTTTAATAGTTGAAAATATTAGTTTGGTTTAAAAATTGTTTTTTATTTGGTGTTTT 13471

Qy        202 GATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTG 261
              |    || |  | ||||  | |  |   || | ||    ||   | | ||  ||||||
Db      13472 TA----TGGTATAATGAAAGTTTGGAACGTTTTTGTTTGTAATTTTGTTAATGTTATTTT 13527

Qy        262 TACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTA 321
              ||   ||  ||   ||  | ||||| | ||||| || | ||    |    || |  | ||
Db      13528 TAGTGAGAAGAGAATAATTTTTTAAATTATAATATTTAAATAATATTTTGATTAGATATA 13587

Qy        322 ATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATA 373
              |  ||  |    || | ||   || ||   ||| || || ||    ||  ||
Db      13588 AAATTTTAGTTTGAAAAATATTTTTGTTTGAAAATTGATTTATAGTATTTTA 13639


RESULT 15
US-10-240-453-107
; Sequence 107, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
;  TITLE OF INVENTION:  Transcription
;  TITLE OF INVENTION:  by Means of Assessing the Methylation Status of Genes Associated
;  TITLE OF INVENTION:  With DNA TRanscription
;  FILE REFERENCE: 5013.1009
;  CURRENT APPLICATION NUMBER: US/10/240,453
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03973
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 107
;   LENGTH: 15732
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-107

  Query Match             8.9%;  Score 46;  DB 6;  Length 15732;
  Best Local Similarity   47.7%;  Pred. No. 5.8;
  Matches  168;  Conservative    0;  Mismatches  180;  Indels    4;  Gaps    1;

Qy         22 TTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAAATCAAAAGCAAAG 81
              ||   ||||  |  ||| || | |    |  ||||   |||  ||||
Db      13292 TTTTGATTACGTTATATTATTTAAGGGAGAAAGTTAAATAGGGATAAAGAAAAATATTTA 13351

Qy         82 TTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAATTTTAAATTGAAT 141
              || ||      |   | || || ||| |   ||||| ||||   ||||   |  |||
Db      13352 GAAGTTTTTTATTATTTTGAATGTGGTTTAATTTTGGTTTGGAATTTTTTTTGTTGTAAT 13411

Qy        142 GTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAGTATGCGGGTTGGT 201
                          || ||| | ||        ||  |||| |  |  |||  ||  |  |
Db      13412 TAATTTTTAATAGTTGAAAATATTAGTTTGGTTTAAAAATTGTTTTTTATTTGGTGTTTT 13471

Qy        202 GATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTG 261
```

```
Db      13472 TA----TGGTATAATGAAAGTTTGGAACGTTTTTGTTTGTAATTTTGTTAATGTTATTTT 13527

Qy        262 TACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTA 321

Db      13528 TAGTGAGAAGAGAATAATTTTTTAAATTATAATATTTAAATAATATTTTGATTAGATATA 13587

Qy        322 ATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATA 373

Db      13588 AAATTTTAGTTTGAAAAATATTTTTGTTTGAAAATTGATTTATAGTATTTTA 13639
```

Search completed: May 12, 2006, 05:34:01
Job time : 340.747 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:52:36 ; Search time 262.392 Seconds
(without alignments)
8004.956 Million cell updates/sec

Title: US-09-423-093B-3_COPY_3361_3875
Perfect score: 515
Sequence: 1 atgtttgtctatagtttaag.....gctggtagtgaaatttgtga 515

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/2/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/2/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	51.8	10.1	624	18	US-11-079-463-2023	Sequence 2023, Ap
2	48.2	9.4	1457619	17	US-11-098-686-8739	Sequence 8739, Ap
3	46.6	9.0	977	12	US-10-301-480-536282	Sequence 536282,
4	46.6	9.0	977	12	US-10-301-480-1149691	Sequence 1149691,
c 5	46	8.9	592	7	US-09-925-065A-491835	Sequence 491835,
c 6	46	8.9	592	7	US-09-925-065A-491836	Sequence 491836,
c 7	43.2	8.4	622	7	US-09-925-065A-780581	Sequence 780581,
8	42.8	8.3	918	12	US-10-301-480-572202	Sequence 572202,
9	42.8	8.3	918	12	US-10-301-480-1185611	Sequence 1185611,
c 10	42.8	8.3	1183	7	US-09-925-065A-722804	Sequence 722804,
c 11	42.4	8.2	666	7	US-09-925-065A-61556	Sequence 61556, A
c 12	42.4	8.2	666	11	US-10-301-480-162794	Sequence 162794,
c 13	42.4	8.2	666	12	US-10-301-480-776203	Sequence 776203,
14	41.4	8.0	613	7	US-09-925-065A-307547	Sequence 307547,
c 15	41.4	8.0	648	7	US-09-925-065A-4562	Sequence 4562, Ap
c 16	41.4	8.0	648	11	US-10-301-480-105799	Sequence 105799,
c 17	41.4	8.0	648	12	US-10-301-480-719208	Sequence 719208,
c 18	41.2	8.0	519	12	US-10-301-480-564593	Sequence 564593,
c 19	41.2	8.0	519	12	US-10-301-480-1178002	Sequence 1178002,
c 20	41	8.0	63609	11	US-10-330-773-514	Sequence 514, App
21	40.6	7.9	521	7	US-09-925-065A-596212	Sequence 596212,
22	40.6	7.9	561	7	US-09-925-065A-592898	Sequence 592898,
23	40.6	7.9	922	12	US-10-301-480-572201	Sequence 572201,
24	40.6	7.9	922	12	US-10-301-480-1185610	Sequence 1185610,
c 25	40.4	7.8	589	12	US-10-301-480-449099	Sequence 449099,
c 26	40.4	7.8	589	12	US-10-301-480-1062508	Sequence 1062508,
c 27	40.4	7.8	603	7	US-09-925-065A-379888	Sequence 379888,
28	40.2	7.8	483	10	US-10-793-626-2371	Sequence 2371, Ap
29	40.2	7.8	495	11	US-10-932-182A-81360	Sequence 81360, A
30	40.2	7.8	495	11	US-10-932-182A-81360	Sequence 81360, A
31	40.2	7.8	613	7	US-09-925-065A-307546	Sequence 307546,
32	40.2	7.8	633	7	US-09-925-065A-664858	Sequence 664858,
c 33	40.2	7.8	648	7	US-09-925-065A-4560	Sequence 4560, Ap
c 34	40.2	7.8	648	11	US-10-301-480-105797	Sequence 105797,
c 35	40.2	7.8	648	12	US-10-301-480-719206	Sequence 719206,
c 36	40.2	7.8	803	12	US-10-301-480-534071	Sequence 534071,
c 37	40.2	7.8	803	12	US-10-301-480-1147480	Sequence 1147480,
c 38	40.2	7.8	998	12	US-10-301-480-548598	Sequence 548598,
c 39	40.2	7.8	998	12	US-10-301-480-1162007	Sequence 1162007,
40	40.2	7.8	3410	10	US-10-793-626-3505	Sequence 3505, Ap
41	40	7.8	570	17	US-11-098-686-9023	Sequence 9023, Ap
42	40	7.8	1758	10	US-10-793-626-2123	Sequence 2123, Ap
c 43	40	7.8	2797	10	US-10-793-626-4369	Sequence 4369, Ap
44	40	7.8	3100	10	US-10-793-626-4063	Sequence 4063, Ap
45	40	7.8	8607	10	US-10-240-708-72	Sequence 72, Appl

ALIGNMENTS

RESULT 1
US-11-079-463-2023
; Sequence 2023, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 2023
; LENGTH: 624
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-2023

Query Match 10.1%; Score 51.8; DB 18; Length 624;
Best Local Similarity 57.9%; Pred. No. 0.3;
Matches 92; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Qy      231 TTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGA 290
            ||||  |   ||| |     ||       ||  | || || |||  | ||    | | |||
Db      201 TTATCCTATAGAACATAAAAAAGGAATCTTAAAAATTGGGGATAATGTAGAAATTAATGA 260

Qy      291 TAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAG 350
            |  ||| |||||| ||     || || | |     |||  ||  || |||||||||| |||
Db      261 TTTTGTGCATATTGCTGCAAGATTATCTGTTCAGATAGGGAATAATGTATTAATTGCTAG 320

Qy      351 CAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAA 389
             ||||| | ||||   |||  |   |||||| |||| ||
Db      321 TAAAGTGTTTATATCTGATATTCAACATGGCTGTTACAA 359
```

```
RESULT 2
US-11-098-686-8739
; Sequence 8739, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
;  APPLICANT: Kapur, Vivek and Gebhart, Connie J.
;  TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
;  TITLE OF INVENTION:  FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
;  FILE REFERENCE: 09531-128001
;  CURRENT APPLICATION NUMBER: US/11/098,686
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: PCT/US03/31318
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: US 60/416,395
;  PRIOR FILING DATE: 2002-10-04
;  NUMBER OF SEQ ID NOS: 11433
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8739
;   LENGTH: 1457619
;   TYPE: DNA
;   ORGANISM: Lawsonia intracellularis
US-11-098-686-8739

  Query Match             9.4%;  Score 48.2;  DB 17;  Length 1457619;
  Best Local Similarity   52.2%;  Pred. No. 5.7;
  Matches  130;  Conservative    0;  Mismatches  118;  Indels    1;  Gaps    1;

Qy        203 ATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGT 262
              |||  |  ||  ||  |||   |||| | |||   |||||||  |    ||     ||||
Db    1135968 ATAGATTATGTTTTTTAGCATAATATTACTATCTAGATGAAATAACATAAAAAGTTTTGA 1136027

Qy        263 ACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAA 322
              |||||  |   | | | | |   | | | | ||  || |   |   |||||||  ||||
Db    1136028 TATTAGGTCAAATTTTGTATATTGATTAAACTTTTTGTA-TAGTAATGTATCATTTTTAT 1136086

Qy        323 TTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATGGCA 382
                |||   |   |  |||  || |  |  ||| |||| | ||  || |     |
Db    1136087 AATTATTGATTAAATTATCCATAGACATAAAAATTTATGTTGGATATAAGCTAGAAATAC 1136146

Qy        383 GTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCAAATAAGCCATTAGGTGATATTG 442
              || |   | ||  | | | |||  ||  |||     |||| ||| ||| |||| |
Db    1136147 AAAATTATTATTTCATGCGGTAGATAATACGACCACGTGTTAAGTCATAAGGAGATAGTT 1136206

Qy        443 CTCCTATTA 451
              || || | |
Db    1136207 CTACTTTAA 1136215


RESULT 3
US-10-301-480-536282
; Sequence 536282, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 536282
;   LENGTH: 977
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-536282

  Query Match             9.0%;  Score 46.6;  DB 12;  Length 977;
  Best Local Similarity   52.3%;  Pred. No. 3.6;
  Matches  103;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        211 TGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGT 270
              |||  |  ||    |||||  |||    ||  | |  |  |    ||| | || ||| |
Db        443 TGGTATATAGGGCAATATATATATATATATATATATATATATATATATATATATATGA 502

Qy        271 GATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAA 330
              |||  |||||  | ||  | | |   || |||| | | | | |   | | ||  | ||| |
Db        503 GATTTTATATATATATATAAGAGATTTTATATATATAGGGGAGATTATATATATATAGGA 562

Qy        331 AACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAA 390
              |   |||||  ||   ||   || | |||||||| |||  ||   ||      |    ||
Db        563 GATTTTATATATATATATACACAAATATATATAGGAGATTTTATATATATATATACACAA 622

Qy        391 GTATGCAGTCCGAAAAT 407
              |||  |    | | ||
Db        623 ATATATATATAGGAGAT 639


RESULT 4
US-10-301-480-1149691
; Sequence 1149691, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1149691
;   LENGTH: 977
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1149691

  Query Match             9.0%;  Score 46.6;  DB 12;  Length 977;
  Best Local Similarity   52.3%;  Pred. No. 3.6;
  Matches  103;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        211 TGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGT 270
              |||  |  ||    |||||  |||    ||  | |  |  |    ||| | || ||| |
Db        443 TGGTATATAGGGCAATATATATATATATATATATATATATATATATATATATATGA 502

Qy        271 GATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAA 330
              |||  |||||  | ||  | | |   || |||| | | | | |   | | ||  | ||| |
Db        503 GATTTTATATATATATATAAGAGATTTTATATATATAGGGGAGATTATATATATATAGGA 562

Qy        331 AACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAA 390
              |   |||||  ||   ||   || | |||||||| |||  ||   ||      |    ||
Db        563 GATTTTATATATATATATACACAAATATATATAGGAGATTTTATATATATATATACACAA 622

Qy        391 GTATGCAGTCCGAAAAT 407
              |||  |    | | ||
Db        623 ATATATATATAGGAGAT 639


RESULT 5
US-09-925-065A-491835/c
; Sequence 491835, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
```

```
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 491835
;   LENGTH: 592
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-491835

  Query Match             8.9%;  Score 46;  DB 7;  Length 592;
  Best Local Similarity   51.5%;  Pred. No. 4.4;
  Matches  106;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        198 TGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTA 257
              || | ||   ||   |  ||    ||| |||| |||     ||  | | |   |   |
Db        534 TGATAATGGGTGTGTGTGTGTGTGTGTATATATATATATATATATATGATAATGTG 475

Qy        258 TTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATG 317
              | | || |||  ||||  | ||| | ||  |||||||||  |||| | |    ||   ||
Db        474 TATATATATATATGATAATGTATATATATATGATAATGTGTATATATATATGATAATGTG 415

Qy        318 TTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCA 377
              | ||  | ||  | |   |  ||  ||   |   || || ||||||   |   |||  ||
Db        414 TGTATATATATGATAATTTGCATATATATATGATAATGTGTATATATATGTATATATACA 355

Qy        378 TGGCAGTTATAAAGTATGCAGTCCGA 403
                  |     | |||||||  |  ||
Db        354 CATTAAAGTCACAGTATGCTCTGAGA 329


RESULT 6
US-09-925-065A-491836/c
; Sequence 491836, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 491836
;   LENGTH: 592
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-491836

  Query Match             8.9%;  Score 46;  DB 7;  Length 592;
  Best Local Similarity   51.5%;  Pred. No. 4.4;
  Matches  106;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        198 TGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTA 257
              || | ||   ||   |  ||    ||| |||| |||     ||  | | |   |   |
Db        534 TGATAATGGGTGTGTGTGTGTGTGTGTATATATATATATATATATATGATAATGTG 475

Qy        258 TTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATG 317
              | | || |||  ||||  | ||| | ||  |||||||||  |||| | |    ||   ||
Db        474 TATATATATATATGATAATGTATATATATATGATAATGTGTATATATATATGATAATGTG 415

Qy        318 TTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCA 377
              | ||  | ||  | |   |  ||  ||   |   || || ||||||   |   |||  ||
Db        414 TGTATATATATGATAATTTGCATATATATATGATAATGTGTATATATATGTATATATACA 355

Qy        378 TGGCAGTTATAAAGTATGCAGTCCGA 403
                  |     | |||||||  |  ||
Db        354 CATTAAAGTCACAGTATGCTCTGAGA 329


RESULT 7
US-09-925-065A-780581/c
; Sequence 780581, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 780581
;   LENGTH: 622
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-780581

  Query Match             8.4%;  Score 43.2;  DB 7;  Length 622;
  Best Local Similarity   45.9%;  Pred. No. 16;
  Matches  147;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy        125 AGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGG 184
              | ||  | |||   ||  |  ||  | | | |  ||| || ||  | || |  |  |  |
Db        413 ATAAAATAAAAACTAACTTACAAAGTCATTGTTAAAATAAAAATTCTTCATATATGTATG 354

Qy        185 GTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAA 244
                  |||  |      |  |  || |   |||     | ||||||  |||    ||  ||
Db        353 TATATATAGGACCCTATTTTTTCTTCCATATTTTCTTTTTATATATATATTTATATATAA 294

Qy        245 AATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTT 304
              | ||        |||   || ||          ||| |  |||| | | ||| |   ||  |
Db        293 ATTTATTTATATATAAATATATTTAAATATTTATGTAATTAATTAAAAATTTATTTAAAT 234

Qy        305 CTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAG 364
                |      | |  |  ||| ||| |||    ||||| ||  |        |  | ||||||
Db        233 AATTATATTTAATTCAAATATTAAAAATATATATAATATATTTATTTTATATRTATATAT 174

Qy        365 GCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCAAATA 424
                |||| ||    | || |||   | | |    || |||  ||     |   || |
Db        173 ATATATATAGTCAACTGAATTTTAACAAAGGTACCAAGATAATTCAAAGAGGAAGTAAAA 114

Qy        425 AGCCATTAGGTGATATTGCT 444
```

```
Db        113 TGTTTTCAACAAATAGTGTT 94
```

RESULT 8
US-10-301-480-572202
; Sequence 572202, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 572202
; LENGTH: 918
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-572202

Query Match 8.3%; Score 42.8; DB 12; Length 918;
Best Local Similarity 48.4%; Pred. No. 21;
Matches 119; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```
Qy        229 AATTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGT 288
Db        460 AACAAAGTAGTTGGATGAATTTATATATATATATATAATATAATATATATTTTTATAT 519

Qy        289 GATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGT 348
Db        520 AATATAATATATATTATATAAAAATATTTTTATATAATATAAAATTATATTTTATATAAT 579

Qy        349 AGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATA 408
Db        580 ATAAAATTATATATATATATACACATGAATGTTTACTACATAGAAAATGTGTAAGTGATA 639

Qy        409 GAACCGCCAGCAAATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGC 468
Db        640 TATCATCTGCCGAGTAACCACTTAAGTATCAACTGGTTTATAAAATAAGAGATTGGGAGG 699

Qy        469 TGGATT 474
Db        700 TGGAAT 705
```

RESULT 9
US-10-301-480-1185611
; Sequence 1185611, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1185611
; LENGTH: 918
; TYPE: DNA
; ORGANISM: Homo sapien

US-10-301-480-1185611

Query Match 8.3%; Score 42.8; DB 12; Length 918;
Best Local Similarity 48.4%; Pred. No. 21;
Matches 119; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```
Qy        229 AATTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGT 288
Db        460 AACAAAGTAGTTGGATGAATTTATATATATATATATAATATAATATATATTTTTATAT 519

Qy        289 GATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGT 348
Db        520 AATATAATATATATTATATAAAAATATTTTTATATAATATAAAATTATATTTTATATAAT 579

Qy        349 AGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATA 408
Db        580 ATAAAATTATATATATATATACACATGAATGTTTACTACATAGAAAATGTGTAAGTGATA 639

Qy        409 GAACCGCCAGCAAATAAGCCATTAGGTGATATTGCTCCTATTAAAATAGGTAATTGCTGC 468
Db        640 TATCATCTGCCGAGTAACCACTTAAGTATCAACTGGTTTATAAAATAAGAGATTGGGAGG 699

Qy        469 TGGATT 474
Db        700 TGGAAT 705
```

RESULT 10
US-09-925-065A-722804/c
; Sequence 722804, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 722804
; LENGTH: 1183
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-722804

Query Match 8.3%; Score 42.8; DB 7; Length 1183;
Best Local Similarity 45.6%; Pred. No. 22;
Matches 146; Conservative 1; Mismatches 173; Indels 0; Gaps 0;

```
Qy        125 AGAATTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGG 184
Db        368 ATAAAATAAAAACTAACTTACAAAGTCATTGTTAAAATAAAAATTCTTCATATATGTATG 309

Qy        185 GTAGTATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAA 244
Db        308 TATATATAGGACCCTATTTTTTCTTCCATATTTTCTTTTTATATATATATTTATATATAA 249

Qy        245 AATTTGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTT 304
Db        248 ATTTATTTATATATAAATATATTTATGTAATTAATTAAAAATTTATTTAAAT 189

Qy        305 CTTGCGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAG 364
```

```
Db        188 AATTATATTTAATTCAAATATTAAAAATATATATAATATATTTATTTTATATATRTATAT 129

Qy        365 GCGATCATAGCCATGGCAGTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCAAATA 424

Db        128 ATATATATAGTCAACTGAATTTTAACAAAGGTACCAAGATAATTCAAAGAGGAAGTAAAA 69

Qy        425 AGCCATTAGGTGATATTGCT 444

Db         68 TGTTTTCAACAAATAGTGTT 49


RESULT 11
US-09-925-065A-61556/c
; Sequence 61556, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 61556
;   LENGTH: 666
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-61556

  Query Match             8.2%;  Score 42.4;  DB 7;  Length 666;
  Best Local Similarity   46.6%;  Pred. No. 24;
  Matches  166;  Conservative    1;  Mismatches  187;  Indels    2;  Gaps    1;

Qy          9 CTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAA 68

Db        430 CTTGAATTAACAATTTGAATACATTCTTAAAGTTTTAAATTTCAGACAGAATAAAATATA 371

Qy         69 ATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAA 128

Db        370 ATTTAAAATAAATATTGTTATTGTCATCAAGGTTTTTTATTAGAAAAGTTTTACTTTTGA 311

Qy        129 TTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAG 188

Db        310 TTCTTTAAACATCAACGATTTTTGTTTTACCGAACTCATTAATATTACAAARTAGTTTTA 251

Qy        189 TATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATT 248

Db        250 GTTAGATATTTTTGTAATTTGTGGGATGAAAACTGAAAGTTTTTTTTCTTTTTAAACTTA 191

Qy        249 TGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTG 308

Db        190 GAATACCAAGTTCAAAAAATATGT--GTCAGGTTATATCAAATAATTTTACATTTTTTTC 133

Qy        309 CGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAG 364

Db        132 AGTATTTGGTTCTTAATTAAAAACAGATTTTATAAAGGCAAACAGTACTTTATTAG 77


RESULT 12
US-10-301-480-162794/c
; Sequence 162794, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 162794
;   LENGTH: 666
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-162794

  Query Match             8.2%;  Score 42.4;  DB 11;  Length 666;
  Best Local Similarity   46.6%;  Pred. No. 24;
  Matches  166;  Conservative    1;  Mismatches  187;  Indels    2;  Gaps    1;

Qy          9 CTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAA 68

Db        430 CTTGAATTAACAATTTGAATACATTCTTAAAGTTTTAAATTTCAGACAGAATAAAATATA 371

Qy         69 ATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAA 128

Db        370 ATTTAAAATAAATATTGTTATTGTCATCAAGGTTTTTTATTAGAAAAGTTTTACTTTTGA 311

Qy        129 TTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAG 188

Db        310 TTCTTTAAACATCAACGATTTTTGTTTTACCGAACTCATTAATATTACAAARTAGTTTTA 251

Qy        189 TATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATT 248

Db        250 GTTAGATATTTTTGTAATTTGTGGGATGAAAACTGAAAGTTTTTTTTCTTTTTAAACTTA 191

Qy        249 TGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTG 308

Db        190 GAATACCAAGTTCAAAAAATATGT--GTCAGGTTATATCAAATAATTTTACATTTTTTTC 133

Qy        309 CGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAG 364

Db        132 AGTATTTGGTTCTTAATTAAAAACAGATTTTATAAAGGCAAACAGTACTTTATTAG 77


RESULT 13
US-10-301-480-776203/c
; Sequence 776203, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 776203
;   LENGTH: 666
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-776203

  Query Match             8.2%;  Score 42.4;  DB 12;  Length 666;
```

```
  Best Local Similarity   46.6%;  Pred. No. 24;
  Matches  166;  Conservative    1;  Mismatches  187;  Indels    2;  Gaps    1;

Qy          9 CTATAGTTTAAGATTAAAATTAAATCTTATCATATCATTATTGAGTAAAGTTAGGCGGAA 68
              ||  | ||  |  |||  |||  | |||||  | | |    || ||   |   ||      |
Db        430 CTTGAATTAACAATTTGAATACATTCTTAAAGTTTTAAATTTCAGACAGAATAAAATATA 371

Qy         69 ATCAAAAGCAAAGTTTCTTGTTCTGCTTAGCGGATATGATTTTAAAATGGTTGGGAAGAA 128
              ||  |||  |||  || || ||  |  | |  |  | | |||  ||||  ||       |
Db        370 ATTTAAAATAAATATTGTTATTGTCATCAAGGTTTTTTATTAGAAAAGTTTTACTTTTGA 311

Qy        129 TTTTAAATTGAATGTCAAACCTTACTCTGCAAAAAATAACACCTCTTCCAAATGGGGTAG 188
              ||  |  |    |    | |   ||  | | |  ||    |  |    | | ||:| |  |
Db        310 TTCTTTAAACATCAACGATTTTTGTTTTACCGAACTCATTAATATTACAAARTAGTTTTA 251

Qy        189 TATGCGGGTTGGTGATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATT 248
                |     ||  ||  |  ||  ||||  || ||| |  |  || |  |  | |||  |
Db        250 GTTAGATATTTTTGTAATTTGTGGGATGAAAACTGAAAGTTTTTTTTCTTTTTAAACTTA 191

Qy        249 TGAACCTTATTTGTACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTG 308
                |  |  | ||  | | |  ||   ||      | ||   |||||  ||    ||| ||
Db        190 GAATACCAAGTTCAAAAAATATGT--GTCAGGTTATATCAAATAATTTTACATTTTTTTC 133

Qy        309 CGTATCATGTTTAATTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAG 364
               ||||   |||     ||| |||  ||| |  |||  |  | ||    ||   |||
Db        132 AGTATTTGGTTCTTAATTAAAAACAGATTTTATAAAGGCAAACAGTACTTTATTAG 77


RESULT 14
US-09-925-065A-307547
; Sequence 307547, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 307547
;   LENGTH: 613
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-307547

  Query Match             8.0%;  Score 41.4;  DB 7;  Length 613;
  Best Local Similarity   51.1%;  Pred. No. 38;
  Matches  119;  Conservative    1;  Mismatches  112;  Indels    1;  Gaps    1;

Qy        203 ATAACTGCTGGATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGT 262
              | | ||||   ||  || ||:|| ||| |  ||       |   |  |||||  |   ||
Db        270 AAATCTGCACAATGAAAACTRTAAATATTGTTGCAAGAAATTGAAGAGAACCCAAAGAGT 329

Qy        263 ACATAGGTGATCGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAA 322
                | ||||  ||      |  |  | ||  |   |  |||||  |    | ||     ||
Db        330 GGAAAGGTATTCTGTGTTCATGTATTGGAAGAATCAATATTGTTACAATTTCCATACTAC 389

Qy        323 TTTTAGAAAACGATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATGGCA 382
              |  ||||| | ||| ||| | ||  |   |  |  | |||  | || | |  | |  ||
Db        390 CTAAAGAAATCTATAGATTCAATGCAATTTATATAAAAATA-CCAATGACATTCTTCACA 448

Qy        383 GTTATAAAGTATGCAGTCCGAAAATAGAACCGCCAGCAAATAAGCCATTAGGT 435
              |  ||| |  | ||| ||| |||||  |   |  | |||| ||| | | | ||
Db        449 GAAATAGAAAAAGCAATCCTAAAATTTATGTGGAACCAAAAAAGACCTAAAGT 501


RESULT 15
US-09-925-065A-4562/c
; Sequence 4562, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4562
;   LENGTH: 648
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-4562

  Query Match             8.0%;  Score 41.4;  DB 7;  Length 648;
  Best Local Similarity   51.4%;  Pred. No. 38;
  Matches   93;  Conservative    1;  Mismatches   87;  Indels    0;  Gaps    0;

Qy        214 ATTGAAGCTGTATATAATTATGGTGATGAAAAATTTGAACCTTATTTGTACATAGGTGAT 273
              ||  ||  | |||   |||||  | ||  | ||| |  |    |||||  || |||    ||
Db        589 ATATAATATATATTATATTATATTTATATATAATATATATTATTTATATATAATATAT 530

Qy        274 CGTATATGTTTAAGTGATAATGTTCATATTTCTTGCGTATCATGTTTAATTTTAGAAAAC 333
              |||||  | ||  | |||    | |||||   |   |||    | ||  ||      |
Db        529 ATTATATTTATATATAATATATTATTATTATATTTATATATAATATATTATATTATA 470

Qy        334 GATATATTAATTGGTAGCAAAGTTTATATAGGCGATCATAGCCATGGCAGTTATAAAGTA 393
              |||||  | |       |  | ||:|||    || |||   ||   || |||||  ||
Db        469 TTTATATATAATATATATTATATATAWATATAATATAATATATATATAAGATATAATATA 410

Qy        394 T 394
              |
Db        409 T 409


Search completed: May 12, 2006, 06:54:56
Job time : 265.392 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 08:31:04 ; Search time 3757.72 Seconds
(without alignments)
15792.702 Million cell updates/sec

Title: US-09-423-093B-3_COPY_3977_5020
Perfect score: 1044
Sequence: 1 atgaatgtttttatcagtat.....ctccgcgtcttggtatttaa 1044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1044	100.0	12441	1	SERFBC2	X61917 S.enterica
	2	1044	100.0	12441	6	BD085556	BD085556 Nucleic a
	3	1044	100.0	12446	6	AX029263	AX029263 Sequence
c	4	65.4	6.3	163443	14	AC006280	AC006280 Plasmodiu
	5	65.4	6.3	196149	14	AC004709	AC004709 Plasmodiu
	6	65.4	6.3	252650	2	AE014847	AE014847 Plasmodiu
	7	63.8	6.1	110000	2	AC116984_2	Continuation (3 of
	8	60.8	5.8	3542	15	YSCMTCG16	L36900 Saccharomyc
c	9	60.8	5.8	192187	2	AC117072	AC117072 Dictyoste
c	10	60.6	5.8	165166	5	CR352263	CR352263 Zebrafish
c	11	60	5.7	202438	14	CR931781	CR931781 Danio rer
	12	59.8	5.7	302156	2	AC116977	AC116977 Dictyoste
	13	59.4	5.7	158863	5	CR392025	CR392025 Zebrafish
	14	59.4	5.7	253924	2	AE014822	AE014822 Plasmodiu
	15	59.2	5.7	98852	5	BX284625	BX284625 Zebrafish
	16	58.8	5.6	204758	14	CR847530	CR847530 Danio rer
	17	58.8	5.6	346939	1	BX842643	BX842643 Mycoplasm
	18	58.6	5.6	36148	2	AC116100	AC116100 Dictyoste
c	19	58.4	5.6	110000	14	PFMAL8P1_03	Continuation (4 of
	20	58.4	5.6	115990	5	BX890614	BX890614 Zebrafish
	21	58.4	5.6	144604	14	BX899179	BX899179 Danio rer
	22	58.4	5.6	145553	5	AL935272	AL935272 Zebrafish
	23	58.2	5.6	4544	2	PFAARPS2	Y08924 P.falciparu
	24	58.2	5.6	183648	2	AC117081	AC117081 Dictyoste
c	25	58	5.6	168659	14	CR855137	CR855137 Danio rer
c	26	58	5.6	182464	5	BX569779	BX569779 Zebrafish
c	27	57.8	5.5	110000	2	AC116984_3	Continuation (4 of
c	28	57.8	5.5	110000	14	PFMAL13_03	Continuation (4 of
c	29	57.8	5.5	110000	14	PFMAL7P1_04	Continuation (5 of
c	30	57.8	5.5	110000	14	PFMAL7P1_09	Continuation (10 o
	31	57.6	5.5	79882	14	BX908758_5	Continuation (6 of
c	32	57.6	5.5	191146	5	BX530070	BX530070 Zebrafish
	33	57.6	5.5	225581	14	BX537105	BX537105 Danio rer
c	34	57.2	5.5	94534	5	AL929250	AL929250 Zebrafish
c	35	57.2	5.5	110786	5	AL845544	AL845544 Zebrafish
	36	57.2	5.5	152951	5	BX927076	BX927076 Zebrafish
c	37	57.2	5.5	182871	2	AC117176	AC117176 Dictyoste
c	38	57.2	5.5	254050	2	PFA929358	AL929358 Plasmodiu
	39	57	5.5	554	2	AF263160	AF263160 Polybioid
c	40	57	5.5	144375	14	CR354586	CR354586 Danio rer
	41	57	5.5	181581	14	AC026621	AC026621 Homo sapi
c	42	57	5.5	181674	8	AC099514	AC099514 Homo sapi
c	43	56.8	5.4	132254	2	AC116330	AC116330 Dictyoste
c	44	56.6	5.4	139489	14	CR376794	CR376794 Danio rer
c	45	56.6	5.4	165427	5	BX537276	BX537276 Zebrafish

ALIGNMENTS

RESULT 1
SERFBC2
LOCUS SERFBC2 12441 bp DNA linear BCT 09-SEP-2004
DEFINITION S.enterica rfbJ gene cluster.
ACCESSION X61917
VERSION X61917.1 GI:47004
KEYWORDS abe gene; abequose synthase; galactosyltransferase; manB gene; manC gene; mannose-1-phosphate guanyltransferase; phosphomannomutase; wbaL gene; wbaP gene; wbaQ gene; wbaR gene; wbaW gene; wbaZ gene; wzx gene; wzy gene.
SOURCE Salmonella enterica
ORGANISM Salmonella enterica
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 12441)
AUTHORS Brown,P.K., Romana,L.K. and Reeves,P.R.
TITLE Molecular analysis of the rfb gene cluster of Salmonella serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B
JOURNAL Mol. Microbiol. 6 (10), 1385-1394 (1992)
PUBMED 1379320
REFERENCE 2
AUTHORS Reeves,P.
TITLE Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale
JOURNAL Trends Genet. 9 (1), 17-22 (1993)
PUBMED 8434412
REFERENCE 3
AUTHORS Liu,D., Haase,A.M., Lindqvist,L., Lindberg,A.A. and Reeves,P.R.
TITLE Glycosyl transferases of O-antigen biosynthesis in Salmonella enterica: identification and characterization of transferase genes of groups B, C2, and E1
JOURNAL J. Bacteriol. 175 (11), 3408-3413 (1993)
PUBMED 7684736
REFERENCE 4 (bases 1 to 12441)
AUTHORS Brown,P.K.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1991) P.K. Brown, University of Sydney, Dept of Microbiology, Maze Crescent, Sydney N S W 2006, AUSTRALIA
FEATURES Location/Qualifiers
source 1. .12441

/organism="Salmonella enterica"
/mol_type="genomic DNA"
/strain="group C2, serovar muenchen M67"
/db_xref="taxon:28901"
gene 138. .1019
/gene="abe"
CDS 138. .1019
/gene="abe"
/codon_start=1
/transl_table=11
/product="abequose synthase"
/protein_id="CAA43918.1"
/db_xref="GI:47005"
/db_xref="GOA:Q00329"
/db_xref="UniProt/Swiss-Prot:Q00329"
/translation="MLDVNKKILMTGATSFVGTHLLHSLIKEGYSIIALKRPITEPTI
INTLIEWLNIQDIEKICQSSMNIHAIVHIATDYGRNRTPISEQYKCNVLLPTRLLELM
PALKTKFFISTDSFFGKYEKHYGYMRSYMASKRHFVELSKIYVEEHPDVCFINLRLEH
VYGERDKAGKIIPYVIKKMKNNEDIDCTIARQKRDFIYIDDVVSAYLKILKEGFNAGH
YDVEVGTGKSIELKEVFEIIKKETHSSSKINYGAVAMRDDEIMESHANTSFLTRLGWS
AEFSIEKGVKKMLSMKE"
CDS 1019. .2359
/note="wzx
Referred to as orf12.6 in reference [2]"
/citation=[2]
/codon_start=1
/transl_table=11
/product="hypothetical protein with 12 predicted
transmembrane segments which possibly functions in export
of O-antigen"
/protein_id="CAA43919.1"
/db_xref="GI:47006"
/db_xref="UniProt/TrEMBL:Q00339"
/translation="MNRIIRMLGVDKAIRYVIFGKIISVLTGLLLIMLISHHLSKDAQ
GYYYTFNSVVALQIIFELGLSTVIIQFASHEMSALKYDYSERDIIGESKNKQRYLSLF
RLAIKWYAVIALLIILIVGPIGYVFFTQKEGLGVPWQGAWLLLTIVTAFNIFLVSVLS
VAEGSGLITDVNKMRMYQSLLAGILAVSLLISGFGLYATSAIAISGTIIFSIFSYKYF
KKIFLQSLKHKNKYTEGGISWVNEIFPMQWRIALSWMSGYFIYFVMTPIAFKYFGAIY
AGQLGMSLTLCNMVMATGLAWISTKYPKWGVMVSNKQLAELSKSFKSAVMQSSFFVLT
GLTGVYISLWLLKLSGSNIGERFLGLQDFFFLSLAIIGNHIVACFATYIRAHKTEKMT
LASCIMALLTITTMLFVAYLEYSRFYMLMYAALTWLYFVPQTYIIFKRFKSSYE"
CDS 2352. .3314
/note="unnamed protein product; wbaR"
/codon_start=1
/transl_table=11
/protein_id="CAA43909.1"
/db_xref="GI:47007"
/db_xref="UniProt/TrEMBL:Q00337"
/translation="MSKKPLLTIAIPTYNRSSCLARLLDSIIQQENYCHDELEVIVCD
NASTDETARIAKSGLDKIRNSTYHLNEENLGMDGNFQKCFELSNGKYLWMIGDDDLIV
KNGISKVFSILKSRPALDMVYVNSAAKTELNYNADVRTSFYTNDVDFISDVKVMFTFI
SGMICKKTDAIVKAVGIFSPQTTGKYLMHLTWQLPLLKQGGEFAVIHNNIIEAEPDNS
GGYHLYKVFSNNLATIFDVFYPREHRVSKRVRASACLFLLNFIGDEDKTKNFATNNYL
RDCDSAFIDLIIYKYGLRFFYLYPKTVPLFRKIKYIIKTVLMRK"
CDS 3361. .3975
/note="unnamed protein product; wbaL
orf14.9 possibly encodes an acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="CAA43910.1"
/db_xref="GI:47008"
/db_xref="UniProt/TrEMBL:Q00480"
/translation="MFVYSLRLKLNLIISLLSKVRRKSKAKFLVLLSGYDFKMVGKNF
KLNVKPYSAKNNTSSKWGSMRVGDNCWIEAVYNYGDEKFEPYLYIGDRICLSDNVHIS
CVSCLILENDILIGSKVYIGDHSHGSYKVCSPKIEPPANKPLGDIAPIKIGNCCWIGD
NAVILAGSEICDGCVIAANSVVKDLKVDKPCLIGGVPAKVIKVF"
CDS 3977. .5020
/note="wbaQ
Referred to as orf15.6 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="rhamnosyl transferase"
/protein_id="CAA43911.1"
/db_xref="GI:47009"
/db_xref="GOA:Q00482"
/db_xref="UniProt/TrEMBL:Q00482"
/translation="MNVFISICIPSYNRAEFLEPLLDSIYNQDYCLKNNDFEVIVCED
KSPQRDEINSIIENYKAKNNKQNLYVNFNEDNLGYDKNLKKCISLTTGKYCMIMGNDD
LLADGALSKIVKVLKANPEIVLATRAYGWFKENPNELCDTVRHLTDDTLFQPGADAIK
FFFRRVGVISGFIVNAEKAKKLSSDLFDGRLYYQMYLAGMLMAEGQGYYFSDVMTLSR
DTEAPDFGNAGTEKGVFTPGGYKPEGRIHMVEGLLLIAKYIEDTTKIDGVYAGIRKDL
ANYFYPYIRDQLDLPLYTYIKMINKFRKMGFSNEKLFYVHAFLGYVLKRRGYDALIKY
IRSKKGGTPRLGI"
gene 5114. .6313
/gene="wzy"
CDS 5114. .6313
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O antigen polymerase"
/protein_id="CAA43912.1"
/db_xref="GI:47010"
/db_xref="GOA:Q00474"
/db_xref="UniProt/Swiss-Prot:Q00474"
/translation="MLPFPPGAILRDVLNVFFVALVLVRFVIDRKKTYFPLVFTIFSW
SAVILWVIALTIFSPDKIQAIMGGRSYILFPAVFIALVILKVSYPQSLNIEKIVCYII
FLMFMVATISIIDVLMNGEFIKLLGYDEHYAGEQLNLINSYDGMVRATGGFSDALNFG
YMLTLGVLLCMECFSQGYKRLLMLIISFVLFIAICMSLTRGAILVAALIYALYIISNR
KMLFCGITLFVIIIPVLAISTNIFDNYTEILIGRFTDSSQASRGSTQGRIDMAINSLN
FLSEHPSGIGLGTQGSGNMLSVKDNRLNTDNYFFWIALETGIIGLIINIIYLASQFYS
STLLNRIYGSHCSNMHYRLYFLFGSIYFISAALSSAPSSSTFSIYYWTVLALIPFLKL
TNRRCTR"
CDS 6313. .7323
/note="wbaW
Referred to as orf17.9 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="second mannosyl transferase"
/protein_id="CAA43913.1"
/db_xref="GI:47011"
/db_xref="GOA:Q00481"
/db_xref="UniProt/TrEMBL:Q00481"
/translation="MNNKKVLMDISWSNKGGIGRFTDEISKLLCDISKEELYRKCASP
LAPLGLAVNIFLRKKTDVVFLPGYIPPLFCSKKFIITIHDLNHLDLNDNSSLFKRLFY
NFIIKRGCRKAYKIFTVSNFSKERIVAWSGVNPNKIVTVYNGVSSLFNADVKPLNLGY
KYLLCVGNRKTHKNEKCVISAFAKADIDPSIKLVFTGNPCNDLEKLIIQHGLSERVKF
FGFVSEKDLPSLYKGSLGLVFPSLYEGFGLPVVEGMACGIPVLTSLTSSLPEVAGDAA
ILVDPLSEDAITKGISRLINDSELRKHLIQKGLLRAKRFNWQNVVSEIEMVLTEACDG
NK"
CDS 7310. .8467
/note="wbaZ
Referred to as orf18.9 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="first mannosyl transferase"
/protein_id="CAA43914.1"
/db_xref="GI:47012"
/db_xref="GOA:Q00483"
/db_xref="UniProt/TrEMBL:Q00483"
/translation="MEINEIKISLVHEWLLSYAGSEQVSSAILHVFPEAKLYSVVDFL
TDEQRRHFLGKYATTTFIQNLPKAKKFYQKYLPLMPLAIEQLDLSDANIIISSAHSVA
KGVISGPDQLHISYVHSPIRYAWDLQHQYLNESNLNKGIKGWLAKWLLHKIRIWDSRT
ANGVDHFIANSQYIARRIKKVYRREASVIYPPVDVDNFEVKNEKQDYYFTASRMVPYK
RIDLIVEAFSKMPEKKLVVIGDGPEMKKIKSKATDNIKLLGYQSFPVLKEYMQSARAF
VFAAEEDFGIIPVEAQACGTPVIAFGKGGALETVRPLGVEEPTGIFFKEQNIASLHEA
VSEFEKNASFFTSQACRKNAEKFSRSRFEQEFKNFVNEKWNLFKTEQIIKR"
gene 8482. .9903
/gene="manC"
CDS 8482. .9903
/gene="manC"
/codon_start=1
/transl_table=11
/product="mannose-1-phosphate guanyltransferase"

```
                     /protein_id="CAA43915.1"
                     /db_xref="GI:47013"
                     /db_xref="GOA:Q00473"
                     /db_xref="UniProt/Swiss-Prot:Q00473"
                     /translation="MSKLIPVIMAGGIGSRLWPLSREEHPKQFLSVDGELSMLQNTIK
                     RLTPLLAGEPLVICNDSHRFLVAEQLRAINKLANNIILEPVGRNTAPAIALAAFCSLQ
                     NVVDEDPLLLVLAADHVIRDEKVFLKAINHAEFFATQGKLVTFGIVPTQAETGYGYIC
                     RGEAIGEDAFSVAEFVEKPDFDTARHYVESEKYYWNSGMFLFRASSYLQELKDLSPDI
                     YQACENAVGSINPDLDFIRIDKEAFAMCPSDSIDYAVMEHTRHAVVVPMNAGWSDVGS
                     WSSLWDISKKDPQRNVLHGDIFAYNSKDNYIYSEKSFISTIGVNNLVIVQTADALLVS
                     DKDSVQDVKKVVDYLKANNRNEHKKHLEVFRPWGKFSVIHSGDNYLVKRITVKPGAKF
                     AAQMHLHRAEHWIVVSGTACITKGEEIFTISENESTFIPANTVHTLKNPATIPLELIE
                     IQSGTYLAEDDIIRLEKHSGYLE"
     gene            9911. .11347
                     /gene="manB"

  Query Match           100.0%;  Score 1044;  DB 1;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 8.5e-157;
  Matches 1044;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3977 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 4036

Qy         61 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4037 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 4096

Qy        121 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4097 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 4156

Qy        181 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4157 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 4216

Qy        241 AAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4217 AAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAAC 4276

Qy        301 GATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4277 GATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCT 4336

Qy        361 GAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4337 GAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGT 4396

Qy        421 GATACTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4397 GATACTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAA 4456

Qy        481 TTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4457 TTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAA 4516

Qy        541 AAACTATCGAGTGATTTATTTGATGGGCGTTTATATTATCAAATGTACCTTGCTGGTATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4517 AAACTATCGAGTGATTTATTTGATGGGCGTTTATATTATCAAATGTACCTTGCTGGTATG 4576

Qy        601 CTAATGGCTGAAGGTCAGGGATACTATTTTAGCGACGTGATGACATTGTCGAGGGATACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4577 CTAATGGCTGAAGGTCAGGGATACTATTTTAGCGACGTGATGACATTGTCGAGGGATACA 4636

Qy        661 GAGGCTCCTGACTTTGGTAACGCTGGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4637 GAGGCTCCTGACTTTGGTAACGCTGGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTAT 4696

Qy        721 AAACCAGAGGGCCGTATACATATGGTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4697 AAACCAGAGGGCCGTATACATATGGTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAA 4756

Qy        781 GATACAACAAAAATTGATGGCGTTTATGCTGGAATTAGAAAAGACTTAGCGAACTATTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4757 GATACAACAAAAATTGATGGCGTTTATGCTGGAATTAGAAAAGACTTAGCGAACTATTTT 4816

Qy        841 TATCCTTATATTCGAGATCAACTCGACTTGCCTCTTTATACTTATATTAAAATGATAAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4817 TATCCTTATATTCGAGATCAACTCGACTTGCCTCTTTATACTTATATTAAAATGATAAAT 4876

Qy        901 AAATTTCGGAAAATGGGATTTTCAAATGAAAAGCTTTTCTATGTGCATGCCTTTTTAGGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4877 AAATTTCGGAAAATGGGATTTTCAAATGAAAAGCTTTTCTATGTGCATGCCTTTTTAGGG 4936

Qy        961 TATGTACTAAAACGGAGGGGCTATGATGCTTTAATTAAATACATTCGTAGCAAAAAAGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4937 TATGTACTAAAACGGAGGGGCTATGATGCTTTAATTAAATACATTCGTAGCAAAAAAGGC 4996

Qy       1021 GGTACTCCGCGTCTTGGTATTTAA 1044
              ||||||||||||||||||||||||
Db       4997 GGTACTCCGCGTCTTGGTATTTAA 5020


RESULT 2
BD085556
LOCUS       BD085556               12441 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085556
VERSION     BD085556.1  GI:22631166
KEYWORDS    JP 2001522248-A/3.
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 12441)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 3 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Salmonella enterica
            PN   JP 2001522248-A/3
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .12441
            FT                   /organism='Salmonella enterica'.
FEATURES             Location/Qualifiers
     source          1. .12441
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match           100.0%;  Score 1044;  DB 6;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 8.5e-157;
  Matches 1044;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3977 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 4036

Qy         61 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4037 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 4096

Qy        121 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4097 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 4156

Qy        181 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 240
```

```
Db       4157 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 4216

Qy        241 AAGAATTTAAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4217 AAGAATTTAAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAAC 4276

Qy        301 GATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4277 GATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCT 4336

Qy        361 GAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4337 GAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGT 4396

Qy        421 GATACTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4397 GATACTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAA 4456

Qy        481 TTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4457 TTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAA 4516

Qy        541 AAACTATCGAGTGATTTATTTGATGGGCGTTTATATTATCAAATGTACCTTGCTGGTATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4517 AAACTATCGAGTGATTTATTTGATGGGCGTTTATATTATCAAATGTACCTTGCTGGTATG 4576

Qy        601 CTAATGGCTGAAGGTCAGGGATACTATTTTAGCGACGTGATGACATTGTCGAGGGATACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4577 CTAATGGCTGAAGGTCAGGGATACTATTTTAGCGACGTGATGACATTGTCGAGGGATACA 4636

Qy        661 GAGGCTCCTGACTTTGGTAACGCTGGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4637 GAGGCTCCTGACTTTGGTAACGCTGGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTAT 4696

Qy        721 AAACCAGAGGGCCGTATACATATGGTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4697 AAACCAGAGGGCCGTATACATATGGTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAA 4756

Qy        781 GATACAACAAAAATTGATGGCGTTTATGCTGGAATTAGAAAAGACTTAGCGAACTATTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4757 GATACAACAAAAATTGATGGCGTTTATGCTGGAATTAGAAAAGACTTAGCGAACTATTTT 4816

Qy        841 TATCCTTATATTCGAGATCAACTCGACTTGCCTCTTTATACTTATATTAAAATGATAAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4817 TATCCTTATATTCGAGATCAACTCGACTTGCCTCTTTATACTTATATTAAAATGATAAAT 4876

Qy        901 AAATTTCGGAAAATGGGATTTTCAAATGAAAAGCTTTTCTATGTGCATGCCTTTTTAGGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4877 AAATTTCGGAAAATGGGATTTTCAAATGAAAAGCTTTTCTATGTGCATGCCTTTTTAGGG 4936

Qy        961 TATGTACTAAAACGGAGGGGCTATGATGCTTTAATTAAATACATTCGTAGCAAAAAAGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4937 TATGTACTAAAACGGAGGGGCTATGATGCTTTAATTAAATACATTCGTAGCAAAAAAGGC 4996

Qy       1021 GGTACTCCGCGTCTTGGTATTTAA 1044
              |||||||||||||||||||||||||
Db       4997 GGTACTCCGCGTCTTGGTATTTAA 5020


RESULT 3
AX029263
LOCUS       AX029263               12446 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 3 from Patent WO9850531.
ACCESSION   AX029263
VERSION     AX029263.1  GI:10190114
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 3 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .12446
                     /organism="Salmonella enterica"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match             100.0%;  Score 1044;  DB 6;  Length 12446;
  Best Local Similarity   100.0%;  Pred. No. 8.5e-157;
  Matches 1044;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3977 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 4036

Qy          61 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4037 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 4096

Qy         121 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4097 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 4156

Qy         181 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4157 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 4216

Qy         241 AAGAATTTAAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAAC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4217 AAGAATTTAAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAAC 4276

Qy         301 GATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4277 GATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCT 4336

Qy         361 GAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGT 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4337 GAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGT 4396

Qy         421 GATACTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4397 GATACTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAA 4456

Qy         481 TTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAA 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4457 TTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAA 4516

Qy         541 AAACTATCGAGTGATTTATTTGATGGGCGTTTATATTATCAAATGTACCTTGCTGGTATG 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4517 AAACTATCGAGTGATTTATTTGATGGGCGTTTATATTATCAAATGTACCTTGCTGGTATG 4576

Qy         601 CTAATGGCTGAAGGTCAGGGATACTATTTTAGCGACGTGATGACATTGTCGAGGGATACA 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4577 CTAATGGCTGAAGGTCAGGGATACTATTTTAGCGACGTGATGACATTGTCGAGGGATACA 4636

Qy         661 GAGGCTCCTGACTTTGGTAACGCTGGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTAT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4637 GAGGCTCCTGACTTTGGTAACGCTGGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTAT 4696

Qy         721 AAACCAGAGGGCCGTATACATATGGTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAA 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4697 AAACCAGAGGGCCGTATACATATGGTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAA 4756

Qy         781 GATACAACAAAAATTGATGGCGTTTATGCTGGAATTAGAAAAGACTTAGCGAACTATTTT 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4757 GATACAACAAAAATTGATGGCGTTTATGCTGGAATTAGAAAAGACTTAGCGAACTATTTT 4816

Qy         841 TATCCTTATATTCGAGATCAACTCGACTTGCCTCTTTATACTTATATTAAAATGATAAAT 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4817 TATCCTTATATTCGAGATCAACTCGACTTGCCTCTTTATACTTATATTAAAATGATAAAT 4876

Qy        901 AAATTTCGGAAAATGGGATTTTCAAATGAAAAGCTTTTCTATGTGCATGCCTTTTTAGGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4877 AAATTTCGGAAAATGGGATTTTCAAATGAAAAGCTTTTCTATGTGCATGCCTTTTTAGGG 4936

Qy        961 TATGTACTAAAACGGAGGGGCTATGATGCTTTAATTAAATACATTCGTAGCAAAAAAGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4937 TATGTACTAAAACGGAGGGGCTATGATGCTTTAATTAAATACATTCGTAGCAAAAAAGGC 4996

Qy       1021 GGTACTCCGCGTCTTGGTATTTAA 1044
              |||||||||||||||||||||||||
Db       4997 GGTACTCCGCGTCTTGGTATTTAA 5020


RESULT 4
AC006280/c
LOCUS       AC006280              163443 bp    DNA     linear   HTG 12-AUG-2000
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   AC006280
VERSION     AC006280.8  GI:9797728
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 163443)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
            Kurdi,O.B., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163443)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1999) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8810453.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   163443: contig of 163443 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163443
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="PFYAC724"
                     /clone="3D7"
ORIGIN

  Query Match             6.3%;  Score 65.4;  DB 14;  Length 163443;
  Best Local Similarity   49.3%;  Pred. No. 0.13;
  Matches  171;  Conservative    0;  Mismatches  176;  Indels    0;  Gaps    0;

Qy         22 TGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATATATAAT 81
              ||||     | | || | ||   |||  | |    |       |||    |||| |||||
Db      18941 TGTAATGAATATGATTACAGGAATGATGTGTGTTCGTTGTGGATGGGGGACATAGATAAT 18882

Qy         82 CAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCCA 141
                |||| ||  |   ||  || || ||  | |     | |  |  |||  ||||   |
Db      18881 TTAGATGATAATAATAATGATGATAATAATAATAATAATAATGATGATAATAATAATAAT 18822

Qy        142 CAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAAT 201
               |    ||||| |  ||   || ||  || ||  ||||    || ||||||||  | |||
Db      18821 AATTATGATGATACTAATAATAATAATGATAATGATAATAATAATAATAATAATGATAAT 18762

Qy        202 CTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCATT 261
               | ||  ||||   ||| |  || ||       || ||||  ||   ||  |    | |
Db      18761 AATAATAATAATAATAATCATAATCATAATAATCATAATAATCATAATAATCATAATAAT 18702

Qy        262 AGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTAGCAGATGGA 321
              | | |   | |  |   |||  | || |     ||  ||     |  | | | |    |
Db      18701 AATATCGATAAACAGACTAATTATAAGAACCAAAAAAATCATTCAAAAAAAAAAAAAAAA 18642

Qy        322 GCGTTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAAATTGT 368
               |||  ||||| || || || |    || | ||||  ||||| | |
Db      18641 AAATTAACAAAAGTATTGGAACTGAATAAAGGTAATAATGAAAATTT 18595


RESULT 5
AC004709
LOCUS       AC004709              196149 bp    DNA     linear   HTG 01-APR-1999
DEFINITION  Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC004709
VERSION     AC004709.3  GI:4558585
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 196149)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B., Conway,A.B.
            and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196149)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Apr 2, 1999 this sequence version replaced gi:4337173.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    47631: contig of 47631 bp in length
            *   47632    47831: gap of unknown length
            *   47832   179129: contig of 131298 bp in length
            *  179130   179329: gap of unknown length
            *  179330   196149: contig of 16820 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196149
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA".
                     /db_xref="taxon:5833"
                     /chromosome="12"
     gap             47632. .47831
                     /estimated_length=unknown
     gap             179130. .179329
                     /estimated_length=unknown
ORIGIN

  Query Match             6.3%;  Score 65.4;  DB 14;  Length 196149;
  Best Local Similarity   49.3%;  Pred. No. 0.13;
  Matches  171;  Conservative    0;  Mismatches  176;  Indels    0;  Gaps    0;

Qy         22 TGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATATATAAT 81
              ||||     | | || | ||   |||  | |    |       |||    |||| |||||
Db      81234 TGTAATGAATATGATTACAGGAATGATGTGTGTTCGTTGTGGATGGGGGACATAGATAAT 81293

Qy         82 CAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCCA 141
              |||| ||  |   ||  || || ||  | |     | |  |  |||  ||||   |
Db      81294 TTAGATGATAATAATAATGATGATAATAATAATAATAATAATGATGATAATAATAATAAT 81353
```

```
Qy        142 CAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAAT 201
              |    ||||| |  ||   || ||  || ||  ||||    || ||||||||  | |||
Db      81354 AATTATGATGATACTAATAATAATAATGATAATGATAATAATAATAATAATAATGATAAT 81413

Qy        202 CTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCATT 261
              | ||  ||||   ||| |  || ||        || ||||  ||   ||  |    | |
Db      81414 AATAATAATAATAATAATCATAATCATAATAATCATAATAATCATAATAATCATAATAAT 81473

Qy        262 AGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTAGCAGATGGA 321
              | | |    | |   |   |||  | ||| |      ||  ||      |  |  | |   |
Db      81474 AATATCGATAAACAGACTAATTATAAGAACCAAAAAAATCATTCAAAAAAAAAAAAAAAA 81533

Qy        322 GCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAAATTGT 368
              |||  ||||| || || || |    || | ||||  ||||| | |
Db      81534 AAATTAACAAAAGTATTGGAACTGAATAAAGGTAATAATGAAAATTT 81580
```

RESULT 6
AE014847
LOCUS AE014847 252650 bp DNA linear INV 11-FEB-2003
DEFINITION Plasmodium falciparum 3D7 chromosome 12, section 4 of 9 of the complete sequence.
ACCESSION AE014847 AE014188
VERSION AE014847.1 GI:23496718
KEYWORDS .
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 252650)
AUTHORS Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W., Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T., James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A., Kyes,S., Chan,M.S., Nene,V., Shallom,S.J., Suh,B., Peterson,J., Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D., Mather,M.W., Vaidya,A.B., Martin,D.M., Fairlamb,A.H., Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I., Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C., Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and Barrell,B.
TITLE Genome sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Nature 419 (6906), 498-511 (2002)
PUBMED 12368864
REFERENCE 2 (bases 1 to 252650)
AUTHORS Hyman,R.W., Fung,E., Conway,A., Kurdi,O., Mao,J., Miranda,M., Nakao,B., Rowley,D., Tamaki,T., Wang,F. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2002) Stanford Genome Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
REFERENCE 3 (bases 1 to 252650)
AUTHORS Hyman,R.W., Fung,E., Conway,A., Kurdi,O., Mao,J., Miranda,M., Nakao,B., Rowley,D., Tamaki,T., Wang,F. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2003) Stanford Genome Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
FEATURES Location/Qualifiers

```
     source          1. .252650
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="12"
     repeat_region   90. .134
                     /rpt_type=tandem
     repeat_region   96. .134
                     /rpt_type=tandem
     repeat_region   586. .669
                     /rpt_type=tandem
     repeat_region   724. .748
                     /rpt_type=tandem
     repeat_region   1220. .1250
                     /rpt_type=tandem
     repeat_region   1331. .1363
                     /rpt_type=tandem
     repeat_region   1440. .1476
                     /rpt_type=tandem
     repeat_region   1758. .1801
                     /rpt_type=tandem
     repeat_region   1907. .1971
                     /rpt_type=tandem
     repeat_region   2150. .2184
                     /rpt_type=tandem
     repeat_region   2301. .2344
                     /rpt_type=tandem
     repeat_region   2484. .2513
                     /rpt_type=tandem
     gene            <2608. .>8601
                     /locus_tag="PFL0930w"
     mRNA            <2608. .>8601
                     /locus_tag="PFL0930w"
     CDS             2608. .8601
                     /locus_tag="PFL0930w"
                     /note="HMMPfam hit to PF00637, 7-fold repeat in Clathrin and VPS, score 2.1e-24"
                     /codon_start=1
                     /product="clathrin heavy chain, putative"
                     /protein_id="AAN36274.1"
                     /db_xref="GI:23496719"
                     /translation="MSQNNPLSVCVADNLINYDIQNESFRLGNVSVEGDKYICVKENV
                     NENTQVVVINLHNKNSTRKHMKAESVIIHPNDPILALKGTIKNMNTIFLQVFNIETKE
                     KICSLNLNEYMNYWKWINNDTIAIVCEKNVYHWNIDIHNTKKNKDNNDNNNNNNNNNN
                     NNNNNNYFNTLTKVFEKAQIFIDNNSQILYYSTDKDMKWCILCGISTQDQGKSIDGYM
                     QLYSCEKKLHQTIEGFIGCFGSFVFDNLDMKPLFCFVEKKKNSTISRLHLMDIYTNKT
                     ETTIPYKIVKEINLINENLNDFPIYISLNTLQGVIYIVTKCSYVYVFDEGTLSLIVKE
                     KISEDNIFICCDSKNGEGIYAVNKKGIIFYITINYFHLINHLKLSNLEFKDKIIKNLC
                     VKYGYPGCDYISAYKKCINDMDFKKASKIICLMKNTKIFEEHSNNIAESILIMNRKNI
                     DIRIIPPLRTQQVLNSFKSFKNTSGQLSPLLLYFSVLLEYDKLNTYESIELVKPVVLQ
                     KKKEYLEKWIKDDKLTCSEELGDLVKVLDLRLALNIYLRCSSHNKILSTYCLLNMFNN
                     VLSYINNFKQINFDYVGIFITIVNYEFPQGNADNKDSAFVGDPNSGSKNETMDFFNDG
                     AASISNDQNNSINKLNKGHINEVAIQYIKFLCENNISFDINKIIDYLLSKKKLQEATS
                     ILLDYLKENKAEHKNLQTKLFEFNLYNNVQVAETLFQMDIFTYYDKNKIAYLCEEKGL
                     YQRALENYTNINDIKRVITKSTCFQKGGNNNNTTTTTTTTSNMGDGHFDMNNVSKGKI
                     SIEWIKNYFSTLSDSVCEELLFDFMKGSKINMEVVISICVQYYDKIGIKKIINKFEEN
                     KNYEGIFYFVSNILNDLPNLIVKGNNNDNLSNNSLSTYNNMNDEDASILLTSDICSTP
                     QYSYETANLKLEDVHYIMFKYIEACVKINNIQELDRICKDKNAKYNPEQIKNFLKDCK
                     LSDPRPLIYVCDIHNYIEELAEYLYKNSLLKYIEVYVIKVNPNNAHKVIGVLLDLDAS
                     EDFLLNLLNNIKNISNIGNLIEIAEKRNRLKLLLPWLESRANEGYENIELHNALAKIY
                     IDLNKDPENFLKNNNFYDKKMIGKYCEDLDPHLAYTAYERSNGECDEELIYITSKNGL
                     FKLQARYLISRQSIELWKVVLDETNKYRKNVIDQVIGSTLIESNNADEITVTVKAFIE
                     KKLSSELIELLEKIVLHNSEFSDNKNLQNLLILTAIKSDSKKVMEYINRLDNFSGPQI
                     AAVAYEYKLREEAFVIYKKFNCYTSAISVLLDKILYNKNQKNKNFHEFSSHDSTVPYY
                     SNNNNKDNIFGNKSDNENDEYLKQFTQNNSYREKYGSYYNNNDISNSNNNNIHLHPEF
                     NNEDKFSSDTLHNYEDIDSTESTCDLSNYEDDLNRAIEFAQKCNNNDVWFILGKAQLK
                     LNKIIDSIDSFVKSNNPEAYKEVIEKCKENNFYEHLITYLNTLREQNSLKDVLVDSEL
                     LYAYAKLKKTNEMNKFINTTNSANLQLIGDRLYKEEEYEVAKILYSNIPNNQKLTACY
                     LKLKEYALAIEAAKKAKSLKTWKEVNFICVKYKQLKYAHTAGLQLIMHADHLDEIIKI
                     YEKKKYINELMNLLENGLNNERAHVGIYTELGILYAKYKPEKLMEFIRNYTNKMNTRK
                     LIDVCENEYLLKEAVYLYISYDEYNLAVDTIIKHSPTAYTADTFMQVIHKVTNSDIIH
                     KVIDFYIEEHPLNLYNLLKILENKIDNNRLVQTLKKSNNLPLIQKYLEDIQAQNITAV
                     NETLNEIYLQNDDYISLRNSIDEYDNFNQTNLLNKLENHKLPEMRRIAALLYKKNKKY
                     KEAINLSKKEKQYKDAIEIARVSKNNVYIEDLINYFIESKNKEAFCACLIVCYDLLKP
                     DYILEIVWTSGFKDQAMLYFIQIISDYTNQIEHMKKQIEDMEKEKKMNKSAPNDYSAM
                     NNQFNYSLNNNLSIMPPQNNFMSSNSFDKYDMFNNNTHF"
     repeat_region   3056. .3105
                     /rpt_type=tandem
     repeat_region   4923. .4948
                     /rpt_type=tandem
     repeat_region   8613. .8642
                     /rpt_type=tandem
     repeat_region   8716. .8742
                     /rpt_type=tandem
     repeat_region   9080. .9130
                     /rpt_type=tandem
     gene            complement(<11613. .>19156)
                     /locus_tag="PFL0935c"
     mRNA            complement(join(<11613. .12941,13715. .>19156))
```

```
                     /locus_tag="PFL0935c"
     CDS             complement(join(11613. .12941,13715. .19156))
                     /locus_tag="PFL0935c"
                     /note="HMMPfam hit to PF03011, Plasmodium falciparum
                     erythrocyte membrane protein (PFEMP), score 1.3e-56"
                     /codon_start=1
                     /product="erythrocyte membrane protein 1 (PfEMP1)"
                     /protein_id="AAN36275.1"
                     /db_xref="GI:23496720"
                     /translation="MVAAAKGGGSSQDAKHVLDEIGKDVYETVEKDAKIYKEALKGNL
                     QHAKGMGERASSNKTCTLVKEYYEHFNGDANSNRYPCKELSGKMGENRFSDTLGGQCT
                     NKKIEGNKNNCGACAPYRRLHLCHHNLESIDTTSTTSDTLLAEVCYAAKFEGETLTTQ
                     HGQHQQTNPGTASQLCTVLARSFADIGDIVRGKDLFLGNDEEKKKRDELEENLQKIFA
                     KIHSGLTSTRGKQNGELQKRYKGDAKKNFYQLREDWWDANRHTVWEAITCGAGESDKY
                     FRQTACGTGTPTHKQCRCDDKANVDPPTYFDYVPQYLRWFEEWAEDFCRKKNKKLKDV
                     KTNCRGDSGTDRYCSGNGYDCTQTIRAIGKLVIGEHCTNCSVWCRLYEKWIDNQKLEF
                     LKQRNKYQTEISGGGASGGASGRSRKTRAATIKYEGYEKKFYEEFKKKGNYGTVDAFL
                     GLLNNEKACKDINEEKEKIDFTKNVEDDKNINKEGTFYHSQYCQPCPGCGMKKTKNGN
                     GWEEKDTNQCNIKLYKPTSSAKPTEIKILKSGEKHDDIKEKIEQFCKTQNGSDGSGGG
                     NSEKKELYDEWKCYQIGELTKDGQEGMDDPDYDRLVETGGGLCILENQKKKEKSDPKS
                     QNEPEQFQKTFNEFFYFWIGRFLNDSMYWRGKVNSCINKSKREKCEEECKKECECFQK
                     WIGKKKEEWDKIKEHFKTQDFGNDALLGPEMKSPSYVLKFVLNIDELFQNIKSGYGNV
                     KELKGINKILDEEKKREEEERADGVVGVASGGDNQNNTTIDKLLDHEDKDATKCKNCQ
                     EPTKPASKPEDLARSGASPDTPREDSPPADLDDEEHDEDDEDEDDDAEEEEEEEKKEE
                     EEEHKEEENEAAPEEKKEEGSSSPEEVEEKAKEAPAGPDACNIVQTLFSSIDKFSDAC
                     TLKYGKTAPTGWKCVTPSGDKAATSSEGKGSSDGGAKDGVGVNGGALQRNKRDLATPS
                     AKSGDTTGGKDGATGKSDGSICVPPRRRRLYIQKLHDWAEKVGDTATQPQVDTPSQSD
                     KLRDAFIESAAVETFFLWHRYKKIKDKEKLEEQQRQRENGELPGLSSSGDGDSNDPQS
                     KLEKGEIPDGFLRQMFYTLADYKDILYSGSNDNTKSSTYNDIINGDKEIQEREKDIKG
                     AISTYFSNSGTTPTPPVKPSVTTPQTWWELHGPHIWNGMICALTYKENGIDKPEVDTA
                     VRAQLWDSGKNTPQNEKYQYTNVKLEEEASGAKINIPDTSGDNTPTLNTPTLKNFVEI
                     PTFFRWLHEWGSDFCRQRKRMLKNVKHNCRNIERGGHEYCGGDGHDCTRDGIQHNNML
                     ADPDCRDCHIQCRKYRKWIEKKFEEFHNQKNTYGKEKQKLNGNSNDGDNTKFFKEIEN
                     RSFDQFLTSLKHCKNNEGDGSDPKNKIDFKEPLKTFGPLEYCKTCPPNKVNCNGPSRR
                     SGGNDQCTAVNGNEWEKIFSENGGNSTTIDVHMIDRRGPYMEKKSQKLENSENPLFKT
                     SRLFKGIREQKWTCKFKDKNTDVCKLDKFDQEVDLNQYTTFKVLLIYWLEDFIEGYYL
                     LKKRKIIEKCTQKGEKTCDDESKNDCACVKKWVEKKTTEWEQIKEHFKNRNQKDGDGN
                     DMKSSVRQLLDPLIYRMDLANGKGKINELKEFLKSYECKCVDNAGNSEKDVVECLLQK
                     LETKAKKCKDPSSGENQKTPCQEYTPPDDDEDLLLEEENTVKAPTFCPKPPKKEEQTE
                     ETCDKAADEKVEETVAQNEENNIPPAGPAPPADSPPAGPATDSGKENVVPEPPAPAAP
                     PSPPIPPLATSTLAWSVGIGFAAFTYFYLKKKTKASVGNLFQILQIPKSDYDIPTLKS
                     KNRYIPYRSGTYKGKTYIYMEGDADEDKYTFMSDTTDITSSESEYEELDINDIYVPGS
                     PKYKTLIEVVLEPSGNNTTASGKNTPSDTQNDIQNDGIPSSKITDNEWNTLKDEFISN
                     MLQNQPKDVPNDYSSGDIPFNTQPNTLYFDNNQEKPFITSIHDRNLYTGEEYSYNVNM
                     VNNDNIPINRDNNPYSGIDLINDSLNSNKVDIYDELLKRKENELFGTEHHPKHTNIYN
                     VAKPARDDPITNQINLFHKWLDRHRDMCEKWDTNNKVDILNQLKEEWENDNSNSGNKT
                     SGNITPTSDIPSGKLSDIPSTNKMLNSDVSIQIHIDKPNQVDDNIYLDTYPDKYTVDN
                     INPVDTHTNPNLVGNINPVDQNSNLTFPSNPNPAYDNIYIDHNNEDLPSKVQIEMSVK
                     NGEMAKEK"
     repeat_region   12602. .12648
                     /rpt_type=tandem
     repeat_region   12960. .13070
                     /rpt_type=tandem
     repeat_region   12966. .13039
                     /rpt_type=tandem
     repeat_region   13048. .13072
                     /rpt_type=tandem
     repeat_region   13469. .13493
                     /rpt_type=tandem
     repeat_region   13562. .13586
                     /rpt_type=tandem
     repeat_region   16563. .16614
                     /rpt_type=tandem
     repeat_region   17915. .17943
                     /rpt_type=tandem
     repeat_region   19242. .19283
                     /rpt_type=tandem
     repeat_region   19334. .19422
                     /rpt_type=tandem
     repeat_region   20141. .20273
                     /rpt_type=tandem
     repeat_region   20940. .21079
                     /rpt_type=tandem
     repeat_region   21067. .21310

  Query Match             6.3%;  Score 65.4;  DB 2;  Length 252650;
  Best Local Similarity   49.3%;  Pred. No. 0.12;
  Matches  171;  Conservative    0;  Mismatches  176;  Indels    0;  Gaps    0;

Qy         22 TGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATATATAAT 81
              ||||     | | || | ||   ||| | |    |       |||    |||| |||||
Db     109117 TGTAATGAATATGATTACAGGAATGATGTGTGTTCGTTGTGGATGGGGGACATAGATAAT 109176

Qy         82 CAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCCA 141
                |||| ||  |   ||  || || ||  | |    | |  |  |||  ||||   |
Db     109177 TTAGATGATAATAATAATGATGATAATAATAATAATAATAATGATGATAATAATAATAAT 109236

Qy        142 CAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAAT 201
               |    ||||| |  ||  || ||  || ||  ||||    || ||||||||  | |||
Db     109237 AATTATGATGATACTAATAATAATAATGATAATGATAATAATAATAATAATGATAAT 109296

Qy        202 CTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCATT 261
               | ||  ||||   ||| |  || ||       || ||||  ||   ||  |    | |
Db     109297 AATAATAATAATAATAATCATAATCATAATAATCATAATAATCATAATAATCATAATAAT 109356

Qy        262 AGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTAGCAGATGGA 321
              | | |    | |   |  |||  | || |       ||  ||     |  | | | |   |
Db     109357 AATATCGATAAACAGACTAATTATAAGAACCAAAAAAATCATTCAAAAAAAAAAAAAAAA 109416

Qy        322 GCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAAATTGT 368
                 ||| ||||| || || || |    || | ||||  ||||| | |
Db     109417 AAATTAACAAAAGTATTGGAACTGAATAAAGGTAATAATGAAAATTT 109463


RESULT 7
AC116984_2
WPCOMMENT
  Sequence split into 6 fragments  LOCUS AC116984 Accession AC116984
     Fragment Name      Begin     End
     AC116984_0             1     110000
     AC116984_1        100001     210000
     AC116984_2        200001     310000
     AC116984_3        300001     410000
     AC116984_4        400001     510000
     AC116984_5        500001     541399
Continuation (3 of 6) of AC116984 from base 200001 (AC116984 Dictyostelium discoideum ch

  Query Match             6.1%;  Score 63.8;  DB 2;  Length 110000;
  Best Local Similarity   50.4%;  Pred. No. 0.27;
  Matches  183;  Conservative    0;  Mismatches  177;  Indels    3;  Gaps    1;

Qy         12 TATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAG 71
              ||| | || || | ||      || ||||      | | ||| |  |  ||     |  |||
Db      39552 TATTAATAATTATTTAAATGGTTTTAATTTTAATAATTTTATGAATTCAAATTTTAATAA 39611

Qy         72 CATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGA 131
               |   ||||| |  || ||  |   || |||||| ||  | |    | |  |  | |  |
Db      39612 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 39671

Qy        132 TAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAA 191
              |||   |    | |   || | |  ||   || ||   | |   ||||    || |||||
Db      39672 TAATAATAATAATAATAATAATAATAATAATAATAATAATAGTAATAATTATAATAATAA 39731

Qy        192 TAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAA 251
              |  | | | | |  ||  ||     |||  | |||||  |   | | | ||  ||  ||
Db      39732 TTTAAATAGTTTAAATAATATAAATAATTTAAATAATCAAAATTTTCAAAATTATCAAAT 39791

Qy        252 AAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTATT 311
              ||| |  | || ||| | |     ||  |||  |  | |     ||  ||||  | ||
Db      39792 AAATTCAAATAATTTAAATAATTTAAATAATTCAAATAATAATTATAATAATGAATTTTT 39851

Qy        312 AGCAGATGGAGCGTTATCAAA---AATAGTGAAAGTTTTGAAGGCTAATCCTGAAATTGT 368
              | ||  |   |  |  |||||   ||| |   |  |||| |||  |||| |  | ||||
Db      39852 ATCATTTTTTGATTATTCAAATCCAATTGAAGATATTTTAAAGAATAATTCAAATATTGA 39911

Qy        369 ATT 371
```

```
                |||
Db      39912 ATT 39914


RESULT 8
YSCMTCG16
LOCUS       YSCMTCG16               3542 bp    DNA     linear   PLN 11-FEB-2002
DEFINITION  Saccharomyces cerevisiae mitochondrion transfer RNA-Ser1 (tRNA-Ser)
            gene and varI gene, complete cds.
ACCESSION   L36900 M62622
VERSION     L36900.1  GI:559276
KEYWORDS    .
SEGMENT     16 of 19
SOURCE      mitochondrion Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 3542)
  AUTHORS   de Zamaroczy,M. and Bernardi,G.
  TITLE     The primary structure of the mitochondrial genome of Saccharomyces
            cerevisiae--a review
  JOURNAL   Gene 47 (2-3), 155-177 (1986)
   PUBMED   3549452
FEATURES             Location/Qualifiers
     source          1. .3542
                     /organism="Saccharomyces cerevisiae"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
     tRNA            1025. .1111
                     /product="tRNA-Ser"
     gene            1767. .2981
                     /gene="varI"
     CDS             1767. .2981
                     /gene="varI"
                     /codon_start=1
                     /transl_table=3
                     /protein_id="AAA67536.1"
                     /db_xref="GI:567922"
                     /translation="MKLKLLNMILSMMNKTNNNNNIIINNTLDSLMNKKLLLKNMLLD
                     MNNKKMNNMKRMLNNNNMNPAGANPVVHRIGPAGNINNKLQHLNNMNNWNTQIYNYNK
                     NMEIMNTMNDKLINKLLYKMMTLKLNNMNINKIIMSKTINQHSLNKLNIKFYYYNNDI
                     NNNNNNNNNNYYMNNMNKLMNIMNNNMNNNLCNILSYYYNKKVTIEPIKLSYIYLNSD
                     IFSKYISLNDMDKYNNGILTNYQRMLNNIMPKLNDHNISMNYINNINNINNNKYNNMI
                     NLLNNNNNNNNNNNINNNNNYNNNNNNYIGNINNIYNNMTIDNIPMDILMYKYLVGWS
                     IKFKGRLSNNNGRTSTTNLLNGTFNNKKYLWSNINNNYKLNYIPSNHNLYNNSNINKN
                     GKYNIKVKLNFI"

ORIGIN

  Query Match             5.8%;  Score 60.8;  DB 15;  Length 3542;
  Best Local Similarity   47.2%;  Pred. No. 2.4;
  Matches  185;  Conservative    0;  Mismatches  207;  Indels    0;  Gaps    0;

Qy          50 TTTTAGAGCCACTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATT 109
               | ||| |    |  ||  ||   || || ||||  ||  || | ||||  ||| | ||
Db        2494 TATTAAATAATATTATGCCTAAATTAAATGATCATAATATTTCTATAAATTATATTAATA 2553

Qy         110 TTGAGGTCATTGTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCG 169
                |      | | ||  | |  ||||||  |     | | |  | |  ||     || ||
Db        2554 ATATTAATAATATTAATAATAATAAATATAATAATATAATTAATTTATTAAATAATAATA 2613

Qy         170 AAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATT 229
               | ||  ||||    || |||||||   | |||  | ||  | ||   ||| |  |||||
Db        2614 ATAATAATAATAATAATAATAATATTAATAATAATAATAATTATAATAATAATAATAATA 2673

Qy         230 TAGGCTATGATAAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGA 289
                  || ||| | |   || |  |  | || | | || |  | |||   |   ||
Db        2674 ATTATATTGGTAATATTAATAATATTTATAATAATATATAACTATTGATAATATTCCTATAG 2733

Qy         290 TCATGGGCAACGATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGA 349
                 ||    |  || |    ||||  | | || |  |   |   | ||  | | |    |
Db        2734 ATATTTTAATATATAAATATTTAGTTGGTTGATCTATTAAATTTAAAGGTAGATTAAGTA 2793

Qy         350 AGGCTAATCCTGAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGA 409
               |  ||||  |  || |         ||    |    |||||  |||||  | ||
Db        2794 ATAATAATGGTAGAACTAGTACACTTAATTTATTAAATGGTACTTTTAATAATAAAAAAT 2853

Qy         410 ATGAGTTATGTGATACTGTTCGTCATTTAACA 441
               ||   | | || ||| |  |  | |||  | |
Db        2854 ATTTATGAAGTAATATTAATAATAATTATAAA 2885


RESULT 9
AC117072/c
LOCUS       AC117072               192187 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 3879572-4071762 strain
            AX4, complete sequence.
ACCESSION   AC117072 AC116964
VERSION     AC117072.3  GI:42733989
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 192187)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 192187)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 192187)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 192187)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 192187)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Feb 21, 2004 this sequence version replaced gi:28830117.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .192187
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="3879572-4071762"
     CDS             5879. .7642
                     /note="GeneID exon scores (in order of location ranges):
                     144.45 - GSCJ_ID dd_03361"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Phosphatidylinositol 3-kinase 3 (EC 2.7.1.137)
                     (PI3-kinase) (PtdIns-3-kinase) (PI3K) (Fragment)"
```

/protein_id="AAO52597.2"
/db_xref="GI:42733990"
/translation="MEVISQNPSIYHIHKSNQNDSHNNNNNNNNNNNNNNNNNNNNNN
YYYNNNNNNNINNNNNNNNNNNNNNNNNNINSIKSPKKISISNSNSSSFYEEGGFDD
MYPSSSSLSSPMMSSSGSINSKSPSPYNVNKQQQQQQQQQQQQQQQQQQQQQQQQLKQ
ISPSTSSTSLMNSPTPSPSTLPNSNSQQQQQQQQQQQTWKSCCGNSRCMLCGNGTPRC
LCMNHPSWTLILRVVFFSLSQLYPDKEFFNLKKEVYVYLNDHWEILSPNQKQSHNWRR
QVQDALSHSRFFRSGRHTLRCNGYWQLKLQCNPWKYINSKQFEALKIDEPGSPQDDFI
NYFPYNVPGEKQVSLQPRTSSSLSNSPSPTPTINSTQTTPTTSPEIAWIISSPKSLSL
TNLKSNQTQTQNQNQIQNQNNQNNQNNNNNNNNNNNNNNNHKINNLIISSSSLNSINIE
GSMDYNDLSSPSSNYHSDDIPSSPLSNSSPSINNAPLNTSQDSNFNFSALTTLSNFSL
TQQPVSIHFPNSLKSSGNNINNTNNNYNNNNNNIKNNINNNNNNSINNNYNNYNIKNN
EKSENNNEKINQMKSILNCTL"
CDS join(10151. .10244,10339. .10554,10594. .11147)
/note="GeneID exon scores (in order of location ranges):
7.64, 12.79, 47.66 - GSCJ_ID dd_02345"
/codon_start=1
/product="similar to Lactococcus lactis (subsp. lactis)
(Streptococcus lactis). Oxidoreductase"
/protein_id="AAS38883.1"
/db_xref="GI:42733991"
/translation="MDPINNIDRVFLVSGTSTGLGLSIVKRLLDNNYKVAAFTRSKLI
VENEIKKYYENNNEKNIDYLNSLLAVQVDITNEESVENGVKETVNKFGHFDVVINNAG
YARAIFDVNYFAILNIVRSTTPYLRKQKSGLILNVSSIVGHLPKGSISSYVASKYAVT
GLTLSLEQELAPFNIKVVLLSPGGFRTIITNKEKFKLVENPIPEYYPNSTPQNSLEAF
IGLMNTSNGDPNKFANAILKLDLIHQQGNKLPSNIFFGSDALEGSSHFFKSLIDESNQ
WKELALSTDRS"
CDS complement(join(11333. .14810,14906. .14919))
/note="GeneID exon scores (in order of location ranges):
171.66, -1.47 - GSCJ_ID dd_03233"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52599.2"
/db_xref="GI:42733992"
/translation="MKNFKYLFKVISKRQSGETYKTITFYSCFISFTICLSLFLEIFK
FSREFFEFPYIIIYVTLFTSTIVSSSDSFNVILKTAFEMMISCIVGDIVAYTIFSIFF
NHLWISIFLYMLFILITSGIFLSSKNDKLNWTIGKRMYFEIFSLLYVVRSPTRSPIIF
LKAGLASIITLLLVTIGCLILNPFFSSELFLRTSVKILHRSHQCNKSLLRQLELKPIN
EFKISPKRKSHKKIIKNIHSRLSNSLSNLPNLSQINNNNNNSISLSNYEIEINNNNNK
ILVLTPKSKKKKEKKLFENEKKKILIENSNLIIEGFREKFIKYNVKYSKNLNMLSTYL
KECDQEFWNKDLVSHFNQLAILLEKNHNMINSMQISIDKDISNPSCHFLLPLIPFINI
IINHSSNIFTIMENQILGNYRISKTKLKIKIFLKKFKKLFNNNNNKNNKDGHDDEDDD
EIDLNKINYKKSLNYSQERILKYFKRIEKVEMEMCKEYSKIEITNDLLHLEEDSQCEI
SRIHFFVNLLTSFTKQQKELSTIVYTMSRCLYRQNRIYHFEMFLLILCWFLNFPKFLY
CKFKIFLKNRKQNKKLKKQRKKQKQEDEYEEDEEDKEEEVENNYLFNFKNRKNEIKLK
IKNNFNYLISIIFCKWKFTIKFTILLSIFSIGYYEILTRSYFILFQNSSWFVVTFLYV
ASPSVGASGFYCCMRFFGTLLGVFSAYVVGVLFSLADSDGTKGLAYIALTFTLVFSIH
FFVRGKPIQSFGNFFILSYATITFPEYTQDHALIITTLLRAFHISLAVFVIMLSSILI
LPYYDHRELEINLLDISIKQLEAFKSIFNHNSHWFDINNNRNIDDLTSISTVSSMIMI
PKEIFQETMRPIFLEIRSLITNQRSLMFNSRFELLFNKQKYKDLKSLLNNISSEYMVL
VGSEFIIDNCQEDGNDSYHNHHHHHHHNSKNYDQTNFLLDVKSLLLPKFNYLIKELDE
CRKSIEEIKSIKIQKKSQRYQNETTTVFRNDNSKQLIKQMYNDYHFLISKHPKAMKYS
FCVHQLGCLIYTFDLFINLMSNITKYIWSITKTILIERKDSNYLYNNNLNSLDNLVTM
SNSSINSSKNINNNNNSTCIDSSTFTSSSSDNNSEDLIIPINRSQFIRHTSIGGTEYN
NMSTLNRVSNCLNIPMN"
CDS join(15424. .15682,15802. .16610)
/note="GeneID exon scores (in order of location ranges):
17.72, 64.18 - GSCJ_ID dd_02343"
/codon_start=1
/product="similar to Arabidopsis thaliana (Mouse-ear
cress). Putative splicing factor"
/protein_id="AAO52600.2"
/db_xref="GI:42733993"
/translation="MENKRKQKDEIVLVNGSNDTTPNKKQKNELQVVSSGGGREIILG
TERTSKLEHPIMQLIGHKGEIYSCKFNSYGTALASGGSDKEIFLWNVYGECINYSVLK
GHKGTILELHWSTDSNEIYTACTDKSIGVWDSNKGELIKRIREHSGVVNSCCPARRGP
PLVASGSDDRSARIFDTRSKGSTHLFQHKYPVTSVCFSDASDQLITGGIDNVIRVWDI
RNQEDPLYTLAGHQDTITSTSVSKDGAYLLSNSMDNSCKIWDIRPYAPPNRNIKTFNG
AQNNFEKNLIKSSWSIDGRRIGCGSSDRQVYIWDTNTKQLQYCLPGHSGTVNEVTFHP
NEPIIASCSSDKTIYLGEIKP"
CDS join(17445. .17449,17576. .17627,17779. .17841,17921. .18315,
18392. .18467)
/note="GeneID exon scores (in order of location ranges):
-1.91, 2.50, 6.66, 36.57, 5.09 - GSCJ_ID dd_02342"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). Ras-related protein RabC"
/protein_id="AAM33191.3"
/db_xref="GI:42733994"
/translation="MEEEILYKIILVGESGVGKSSILVRFTDNTFSQHFAPTLGVDFN
VKTIRNKETGQTVKLQLWDTAGQERFKSITQTFYRGSHGVIVVYDVTDPKSFERCKNW
VEDINQYTQDGMIIILVGNKSDMVAQRKVTFEQGQEMAEQLKTKFLEVSAKENNGVTQ
VFDLLVQDIEATMKNSKVAQNQLNLNSEVGQKRGCC"
CDS join(19787. .19807,19967. .20046,20176. .20472,20545. .20902,
20995. .22992)
/note="GeneID exon scores (in order of location ranges):
2.08, 8.54, 23.29, 38.31, 196.23 - GSCJ_ID dd_01931"
/codon_start=1
/product="similar to Plasmodium falciparum. Hypothetical
protein"
/protein_id="AAM33151.3"
/db_xref="GI:42733995"
/translation="MDVDKQMLQWRKRAWLNSNQLYGELTDIIIKDTGSFANLIDNSE
VDSFSKNIVNIFTTANKTLPLIKDLIYNEFNTKVLAEGEGSILRGNSIVNKIEGAYVR
LIGANYLRFVLSDLVTKVVLDSDLKLEIDPRKLNDYYEDKIEFEKKVAEVELRDNQKS
LHNIAQQFLDRITDPSVVVEMPREIRAIADYTAESALRYAPESLAPLVGGFIMLRFFS
PAIVTPEYSKLLSSEVAPSKRAKRNLVLLAKVLQNASNGVLFGGKEEFMTCMNDFIIG
NKEKMSAYFNLICKDPLNQNNKWSDLEGVKSATTTTTILTPTNTTTSSNGSGGTTWAK
PIVGGSKWLATTPSGNTPSPAISNASSAHNGKSNNTTNNNNNNNNNNNNNNNNNNNN
NNSNKTTTTATTISSNSWDVYIKNVQIQDLFDLHRIFDSYKEKIVNKIINTNNNNANN
ANNNNTNNINNKTAIRVIEILKELGPSPKTKTERKKKETDAAEEPTTSLQNGASMGSA
FDECTHMLERSRFLFQGPNDKNDRSVFYLIVNRVKPEVFDNVNPLIAHIFKVMDICVN
SPYTLVVDMSWAHISNDLKKAIFTHLPKLAEIFSRKYKKNIDKIFIVHPSAYTRAVIY
FMSAFTSRKLKRKIHDIYNWKDLSQYIDIENIALPETSKDFITKSYRVVKVNSKGKRQ
ERLIKFTSNSLLNIDPKTHRIQNEKRINEIDEITSRLGSIEINMKLSGESKKSGSILG
TKIGFLSINNNNNNNNDDSNGNHDINESNSRKYICNHELERDHILQDIFETGFKMGLQ
SKSTKSLPTEYKVIKVNNVGKHQERIFKLTIDSLLNLDQQRIKSENSFAGIEEVILDP
HDDDIVWLKFKSEPSRRKIICTVKGEGKLLLEVLTDAINKYQTTIDIQEGALKLDAEE
GFL"
CDS complement(join(23340. .23728,24311. .24317))
/note="GeneID exon scores (in order of location ranges):
69.55, -1.92 - GSCJ_ID dd_01928"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAS38884.1"
/db_xref="GI:42733996"
/translation="MTTTTRHQEHEHKVGFPRALVAENEEHHKVGVPRALVAENEEHH
KVGVPRALVAENEEHHKVGVLRALVAENEEHEHEHKVGVPRALEAENENKVGVPKALV
AETTETVDEKKPPIGTIHVLKSDFLPMLR"
CDS complement(24405. .25106)
/note="GeneID exon scores (in order of location ranges):
60.14 - GSCJ_ID dd_01925"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAS38885.1"
/db_xref="GI:42733997"
/translation="MKLLLSLLVLVNILILNVLAQHGDHHHSDVHFTQFLIQGNGCSA
GGELLHVNECFSNVCGVSGALFSEVDYVHMEYLLSLTNDTSCKNTNLSTVFECSDENV
AVSLGNGYSVTCYDDEYVVPKYTKFQISGPGCSNDGETISNVDECSTICGSNNKVVEK
TETSFNTFSLNSYSDATTTKTCSTLKNSQDFICLPDSQKVSVGNYSIICESSSSLVSA
SIFVTVLILSIVLLI"
CDS join(25928. .25976,26103. .27193,27260. .27508)
/note="GeneID exon scores (in order of location ranges):
-0.51, 97.70, 27.06 - GSCJ_ID dd_01918"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). Gamma tubulin"
/protein_id="AAS38886.1"

Query Match 5.8%; Score 60.8; DB 2; Length 192187;
Best Local Similarity 51.1%; Pred. No. 0.68;
Matches 143; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy 68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
||| | ||||| | || || | || |||||| || | | | | | | |
Db 165116 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATTAATAATAATA 165057

```
Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |   || | |  ||   || ||   | ||  ||||      || |
Db     165056 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 164997

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| |  |||||       || |||| |||
Db     164996 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 164937

Qy        248 TAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATC 307
              || ||    | ||| |   |         ||  | |     || |||| ||| | | |
Db     164936 ATAATAATAATAATAATAATAATGGTAAAAATAAATATGAAATTATGGTCAATGGTAACA 164877

Qy        308 TATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTT 347
              ||   |    | | | || | ||| | | || ||| |||
Db     164876 TAAATGATATAGAATCATTTTTAAAGAAATTGGAAGATTT 164837
```

RESULT 10
CR352263/c
LOCUS CR352263 165166 bp DNA linear VRT 19-MAR-2005
DEFINITION Zebrafish DNA sequence from clone CH211-120M16 in linkage group 3, complete sequence.
ACCESSION CR352263
VERSION CR352263.9 GI:61657062
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 165166)
AUTHORS Lovell,J.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Mar 19, 2005 this sequence version replaced gi:54695469.
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.
Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml
CH211-120M16 is from a CHORI-211 BAC library
VECTOR: pTARBAC2.1.
FEATURES Location/Qualifiers
source 1. .165166
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-120M16"
/clone_lib="CHORI-211"
ORIGIN

```
  Query Match             5.8%;  Score 60.6;  DB 5;  Length 165166;
  Best Local Similarity   57.8%;  Pred. No. 0.77;
  Matches  108;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||  | | |||||  || || ||  | | || ||||||  ||  |   |||| |  ||
Db      10869 ATAATAAAGATAATTAAAATAATAATATTAATAATAATAATATGAATAGTATTGATGATG 10810

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |   || | |  ||   || ||   | ||  ||||      || |
Db      10809 ATAATAATAATAATAATAATAATAACAATATTAATATTATTATTAATGATAATAACAATA 10750

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              ||||||  | | |||  ||||  || | | |||  |  |||||  ||    ||  |||| |||
Db      10749 ATAATATAAATAATAATTATAATATTATAAATAATAATAATATAAATAATAATAATAATA 10690

Qy        248 TAAAAAA 254
              || ||
Db      10689 ATAATAA 10683
```

RESULT 11
CR931781/c
LOCUS CR931781 202438 bp DNA linear HTG 10-JUN-2005
DEFINITION Danio rerio chromosome 11 clone DKEYP-71H2, WORKING DRAFT SEQUENCE, 6 unordered pieces.
ACCESSION CR931781
VERSION CR931781.4 GI:67509332
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 202438)
AUTHORS McLaren,S.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT On Jun 10, 2005 this sequence version replaced gi:56605455.
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
---------- Project Information
Center project name: zKp71H2
---------- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 200501 bases at least Q40
Consensus quality: 200988 bases at least Q30
Consensus quality: 201344 bases at least Q20
Insert size: 201938; sum-of-contigs
Insert size: 197200; 4.2% error; agarose-fp
Quality coverage: 7.98x in Q20 bases; sum-of-contigs Quality coverage: 8.38x in Q20 bases; agarose-fp
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9378: contig of 9378 bp in length
            *    9379     9478: gap of 100 bp
            *    9479    17661: contig of 8183 bp in length
            *   17662    17761: gap of 100 bp
            *   17762    63017: contig of 45256 bp in length
            *   63018    63117: gap of 100 bp
            *   63118   123323: contig of 60206 bp in length
            *  123324   123423: gap of 100 bp
            *  123424   169916: contig of 46493 bp in length
            *  169917   170016: gap of 100 bp
            *  170017   202438: contig of 32422 bp in length.
FEATURES             Location/Qualifiers
     source          1. .202438
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="11"
                     /clone="DKEYP-71H2"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .9378
                     /note="assembly_fragment:00040
                     fragment_chain:1"
     misc_feature    9479. .17661
                     /note="assembly_fragment:00115
                     fragment_chain:1"
     misc_feature    17762. .63017
                     /note="assembly_fragment:00727
                     fragment_chain:1"
     misc_feature    63118. .123323
                     /note="assembly_fragment:01283
                     fragment_chain:1"
     misc_feature    123424. .169916
                     /note="assembly_fragment:02017
                     fragment_chain:1"
     misc_feature    170017. .202438
                     /note="assembly_fragment:00213.0"
ORIGIN

  Query Match             5.7%;  Score 60;  DB 14;  Length 202438;
  Best Local Similarity   55.2%;  Pred. No. 0.9;
  Matches  117;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy         69 TAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGA 128
              ||||||| || |||  ||  |  | | || |||||| ||  | |    | |  |  | |
Db      51445 TAGCATAATTATTCACTATATTATTATTAATAATAATAATAATAATAATAATAATAATAA 51386

Qy        129 AGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAA 188
                ||||   |    |    |||||     |   |  ||   | ||  ||||| |||| ||
Db      51385 TAATAATAATAATGATGATGATGATGATGATGATAATAATAATAATAAATCAAATAA 51326

Qy        189 TAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTT 248
              ||||||  | ||  || ||  ||||   ||| |  |||||       || |||| |||||
Db      51325 TAATAATAATAACATTAATAATAATAATAATAATAATAATAATAATAATAATAATTT 51266

Qy        249 AAAAAAATGCATTAGTTTGACGACAGGTAAAT 280
              |  |||||  |  | ||  |  | |   ||||
Db      51265 ATTAAAATAAACAATTTAAAAAATAAACAAAT 51234


RESULT 12
AC116977
LOCUS       AC116977              302156 bp    DNA     linear   INV 25-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 5515173-5817331 strain
            AX4, complete sequence.
ACCESSION   AC116977
VERSION     AC116977.2  GI:28828573
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 302156)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 302156)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 302156)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 302156)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 302156)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Mar 4, 2003 this sequence version replaced gi:20042932.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .302156
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="5515173-5817331"
     CDS             join(<1385. .3437,3519. .4001)
                     /note="GeneID exon scores (in order of location ranges):
                     178.05, 30.36 - GSCJ_ID dd_01590"
                     /codon_start=2
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Putative transcription factor"
                     /protein_id="AAO51288.1"
                     /db_xref="GI:28828685"
                     /translation="NNNNNFDPITSLTNNNNNKIISPLNSNRQQPNSNTNNNSNNKNN
                     NKSNLLDIILDNNNLENNIDNNNNNNNNNNNNNNNNNNNNSKSIINLIDNNNNNNKNN
                     NNNNNNNNILKIPIDYTPVDKNTESIFKGFSPFSFKQQQPIDTSDQQQKETKKLQLNI
                     NSNNKNNNNIINNSNNNNNINNIESNNNNLNSASTSPIPTPTSLTAVSVPSSSPLPSP
                     TSPNTILQSPTTTTTTTTTTSNSSVSPQPTQINKTDKYLLPMSAPRSCEFCGSSQTPT
                     WRRGPSGKGSLCNACGIKWRLKGKDGIFKPSQKQQNRQKPIMSAQKQPKQQQNQPQQQ
                     QPQQPQQPQQPQQQNQQQNQQQSIQSQPQQRQNNHLILNQQTINQHLQMLSKQHQQQA
                     SQSMFQQQQLHNNQHSNDSNYNSYNNNNNNNNNNYSQHEQLPPISSYNSMNNNNNSNN
                     SNNNNNNNNYSQIPQYGSDDNVSVPNQSPDSYKPSKKRKPEFIHTDTSGNPIEFEKGY
                     YCKYCKKTWPQSSFRNSQQFGAHCSNCSRKPKGELEELISIKKKLSKKKSKEGNYYQE
                     EDGAVPYWDMNRYNNNNNNNNGINNNNNNNNNNNNNNSNNNGSYNNNNSNNNNNNRI
                     EPIVNNNINNFNSINNNNSNNNNNNNSTSQNAPLLQRLIHVVENQLLEPNELSSIRED
                     IDIMKNDLILRKKKRIAQLNDLKQKANNDVQILKANIEKGSNQRSNESKSFITDLKQE
                     IDKRLNQYTHIYIPPDHHLVPSSTNYLSIDSLQNSSSSNNYGTIHNSNNNNNNNNNNN
                     KRSLVNLSPELIHSPLKKPMVGGYRDEQGMKHFSIDGDNLSLPSIS"
     CDS             complement(4147. .4578)
                     /note="GeneID exon scores (in order of location ranges):
```

```
                21.45 - GSCJ_ID dd_01589"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO51177.1"
                /db_xref="GI:28828574"
                /translation="MDSNSTISLMTDSKAELVAYYSGRVVNKSSKDIIVIAVPIIAGT
                IFFTIICLGILFRKRISSIIQKNKSNKEKYNINDIKIPQSPSTIITTNSKNNNSCNYL
                NSSDISMVLKNSNDNNDNNIENAFEIIEIPHLDIELIKFKK"
CDS             complement(4816. .9204)
                /note="GeneID exon scores (in order of location ranges):
                317.53 - GSCJ_ID dd_01588"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). Hypothetical 12.5 kDa protein (Fragment)"
                /protein_id="AAO51178.1"
                /db_xref="GI:28828575"
                /translation="MLSLKMINNYNCERLNKSSSEEVLFWKVFGNVYLLKLIFHHIHT
                TEWVVYDAYKKINVQNRWKFKDVYSLDWMILKKQYPLLINKIKFGDFILFDTISIKVL
                FSAIPKLKKQPPPPPPLTTTPSSNSTTKTLTLKAVKEKEKEEKRKEIEKEIENEKEIE
                TLKYIIKLLLEKRRLEFEICDLTQLAVKTNCLDTLSLLLCEQYSTIVYPSTLEWAIEN
                CDCSIVKQLLTYNNKQLLTEDIKKKAFFLASINTHYQNEMINLILDESLYKLPPPTPT
                PTPPPPPPTSSSPQHQKQIDQLSKNQNKKIKPDSNSDYHNYGNDDNENDNDDDDDDDY
                IFGDDYEEDNRDILDSYYDDYFGNHETLSEIERERVTEYPYIKFEKFLTITSIDTMSK
                LMDLKVVKPIARQHYQMDLFKFEFSLNQLLVYLRIYLSLFKKDYYFYSKLYETLEIKI
                NNEIINNEKLDQSAKLKELRRLVVIHSRQETFYKQYILHYEESIDEFHRDFPYDRIDS
                SNLDSDCKKFLALLRYMGEFVAHKMPNRHKYFLEDLKKMKSSTSNDRWCPLLFKAIEC
                VPGQLENFNITHKSIIDWIKNTNSIVINTAVCTSFTMVSYSLIELKHYAITEAFTNLD
                MEDNKKSLSYFCKKNWFVKFYLKEMEFCNVERLREIEDFLTKNCLENPSSPININVDS
                YGSFDFKIDQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQKQQQKQPNQNIVDDQFDSDV
                FEKKIVALIEYLNKTNFKFIQHGMHAFLKKVIEHLPREISSTLVVNLNPLKFYLNSWL
                TASIITLNLKMVEFLINNLEAIEFEEQDFTRIQSVERNLFNSSLNLYHNHDEILVFDF
                VMSIFKTIVNKWRCTVNNDHQKEGGSGSGGSSSGTIKSLPESCIVILNYFFEMIIQTKY
                ITIEQVRMAHQFLCLNSIPIKHSLFHAFYYLKIRSAKFIKYILSRPGIETFFDVTMDS
                RYNSTREYFTERCFDIRDHLKYEMFTFDYVLGGESSSSGGASSTAASFDEILQNLFDQ
                LSNDICFKENIVHNFYQLSKNADLSFCLDIRRVDLFFKHLDIIQSHLKDTTSEDYVLK
                SNDPFPIQEYSYNSPMTTILPTSQLFSGGREFPKPYLPFCERNTIQYDHCIPKPLTII
                NSFDYSEKVLDLIFMVITDLELIQKFLEYNFHVIQFEDTLFSKSLHWSRFDISTFLLA
                NFDIQFFSFPRFALDTHYSFYQEAVFILNNHFDKVRYPNKFKYDLNVFNDTTFPIAFL
                SVLIDFLSKRKNSNHYEKFMNIETRKDALIKLLRIQRDTYGASNVDDDMDADDINFET
                SFQKEKQCRAILIQPETQSLLKSIYDDFKIELDKKREKEKLTGEAEYLDSDDNLFLSA
                FIVGDIKLCDLLLCHYYHPIQFKITKSAITKAIKDQKIHIIKYYYNGHHLNLLNNNIK
                NDKTLSDHLLKELLNHNDYKFDLNWL"
CDS             complement(join(9977. .10749,10855. .11323,11395. .11661,
                11921. .12127))
                /note="GeneID exon scores (in order of location ranges):
                79.65, 53.78, 36.02, 15.96 - GSCJ_ID dd_01587"
                /codon_start=1
                /product="similar to Homo sapiens (Human).
                Phenylalanyl-tRNA synthetase beta-subunit (Fragment)"
                /protein_id="AAO51179.1"
                /db_xref="GI:28828576"
                /translation="MKKNETGVEDLTVSDDVIYKIDVSANRYDLLCLEGIARALNVYN
                HKASIPKYQIVPPKNSHEKLYISKEVESVRPVIVAGILRDITFTQESYDSFIDLQEKL
                HANICKKRSLVSIGTHDLDTLSGPFYYKALAPKDIKFVPLSQTKEYNAEELFKFYDES
                SSHLKKFLPIIKDSPVYPVIYDSKNVVCSLPPIINGEHSKIKLSTKNVFIEVTANDRT
                KANIVLNTMLTMFSEYCKQPFTMEQVEVIDADGKSTGLYPQIQEKQINAQVDYINKSA
                GINITPNDMVTLLKRMSLQSKLSDDEKSIIVDVPVTRSDIMHACDIMEDVAIGYGYDN
                LKKEIPNCNTIGRVQPINKLSELLANEIALAGFTEIMTFVLCQNRDNFTALNKADDGS
                SVKISNAVSEEFTEVRTNLVSTLLKSVSANKAAPLPLKMFEISDVSIKGSLGNKDLSD
                PNSNNSDVGAYNKRMLGAIYCNQSAKIEVIHGLLDRIMLVLDIKLDATRSSNKGYYLE
                LSNDKLFLPGTGINVIVNGKRVGHMGIVHPLVLKNYSCSFPCTILELELTIDTMAHNV
                LLREN"
CDS             13521. .13895
                /note="GeneID exon scores (in order of location ranges):
                1.38 - GSCJ_ID dd_01586"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO51180.1"
                /db_xref="GI:28828577"
                /translation="MNQQELIEEINKIESCSLVFERFFGYLPNDIKERKQYILKFNKT
                LKDEKRKLKCKKTFKGEKKNKKQTFLKHHQSDDHSENKVYKSKKLEKKIQQLNKKKQL
                IDTKINFLKEIKESNKINKSKI"
CDS             complement(join(13981. .17874,17997. .18239))
                /note="GeneID exon scores (in order of location ranges):
                419.33, 16.26 - GSCJ_ID dd_01585"
                /codon_start=1
                /product="similar to putative DEAD/DEAH box helicase
                [Schizosaccharomyces pombe]"
                /protein_id="AAO51181.1"
                /db_xref="GI:28828578"
                /translation="MIELNIDSFLEELNNLENITSTNTTPVNKSNSSIKNNSNNTNVE
                SFDLPNWYVNPTGLSAQPEITPIAPQSSIISPTIPNNCIQLFNFSDKEEILSDIRREY
                LDFSIESHKLNDDDEQKRQSYLIWPRNIDTDSWDFKLLAERTLSIPCSTVELVIDKIQ
                YSISITNYHEVFLNTEKTSSNSTSLDRPWSNSISKDNFHRGNTSNLPFLPGGIKSPNN
                NSTNSTNNNNQTPIDWVQFWHSKSLLVKPPGLDNGILDFNENINKQDNKQQQQQQQQQ
                QQQQQQQQQQQEKEKEEEEEVNLRYTLKDLEREIEEERKREEAEEAEVYEDDDQEDDL
                DIEEEELDELDDIEKTKEEIDDIIKSPPQPTSSSLSTSSSSSSTTLKDKNKIEIEIEE
                ETKRKEEESKKEKQWAFLEQKEITSPLSDLITNPAIEYPFDLDSFQKQAIVHMEQGES
                VFITAHTSAGKTVIAEYAIAMAAKNMTRAIYTSPIKALSNQKFRDFKNTFNDVGLITG
                DVSISPSSSCLVLTTEILRSMLYKGADLIRDIEWVIFDEVHYLNDLERGVVWEEVIIM
                LPPYVKMVFLSATVSNPLEFAQWIGRTKQLPIYVIGTTKRPIPLEHYIHTPSNELFKI
                VDSNRNFLPSGYNSAYNSLYKTDNNKGGGGGGRGGHNQQQQQRGGSGNNISGWSKLIN
                TLKDKQQLPVIIFSFSKNKCQEYAGSLGQSVNLTQGNEKSQIRVFIEQSLGRLCEDDK
                SLPQILQMKELLERGIGVHHGGLLPIVKELVEILFSKSLVKVLFATETFAMGVNMPAK
                TVVYSSTRKHDGITFRDLIPGEYTQMSGRAGRRGLDKVGTVIITYWKDIPEQATIESM
                ILGTPSRLNSQFRLTYNMILNLLRVPDFKVEDMIKRSFSEFSSQKELPGIEKQIEKLQ
                QQQKQLAQVDCILGEPDIDHYYKLFSQAKKYNQQIQNIILNLPNDNHLSIGRVVVLST
                SSNIEKEVNDDDDDDDDNNDDDEIFENEKYTIGLILECKSTLIKQYTNTKVDRVFKIF
                SLKLKIENDIVLCVGNEIITSSGGKEIMKVCNEKLQLKQSDLKSILAGAESGGESEIL
                INSLEQQLLRLIEEYPLPLGPKSIDPIKQLKLKDVDFVSTYDHLQSIEKLIPESKCHK
                CPRLHDHYEQTEKRYQLQYAIRDAKYTASDENLKLMPQFNIRLDILHELGYIDDENTV
                TLKGRVSREINTCEDLVITELIFENAFINLEPSEVVSVLSCLIFQEKDAVQPSLTPRL
                EEAKQNLIKTAEKTYKVESDKGLDVVPDDKLETTLKFGLMQVVYEWARGTPFNDICTL
                TNVLEGSIVRAITRIGETCQEVRNAARVIGDTKLLQKMEEAMRLIKRDIVFTSSLYVN
                "
CDS             complement(18767. .19162)
                /note="GeneID exon scores (in order of location ranges):
                41.07 - GSCJ_ID dd_02896"
                /codon_start=1
                /product="similar to Involved in global regulation of
                transcription; Snf5p [Saccharomyces cerevisiae]"
                /protein_id="AAO51182.1"
                /db_xref="GI:28828579"
                /translation="MEIINYLPPAACENCPQHKYCTRCYFQQLQFMSALTIHQKEISQ
                PKKRKNTGHDEPNSNFYESDIESDDDDEQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
                QQQQQQQESESRTVTGIQIEFDDEIEFYH"
CDS             complement(join(20381. .22148,22235. .23160,23330. .23685,
                24790. .24793))
                /note="GeneID exon scores (in order of location ranges):
                165.28, 71.34, 7.11, -6.65 - GSCJ_ID dd_02895"
```

```
Query Match             5.7%;  Score 59.8;  DB 2;  Length 302156;
Best Local Similarity   50.5%;  Pred. No. 0.85;
Matches  145;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||  |   ||||| || || ||  | |||| ||||   ||  | |  | | |     |
Db      75748 ATAATAATAATAATAAAAATAATAATATAAATAATATAAATAATAATATAAATAACAATA 75807

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
               | ||||   |  | | |   ||   |  ||    | ||   | ||  ||||    || |
Db      75808 TAAATAATAATATAAATAACAATATAAATAATAAAAATAATAATAATAATAATAATA 75867

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| |  |||||        || | || | ||
Db      75868 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAAAATAGTT 75927

Qy        248 TAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATC 307
              || |||     |  |   | | | |   ||||  |  || ||   |   ||  ||  ||
Db      75928 TAGAAATTATTAAAAAAATTAAAAAATCAAAATCATTTATAATTTGGAAAAAACATAGTC 75987

Qy        308 TATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCT 354
               |  || | ||  |   |||  ||||    | | | | || |  |||
Db      75988 AAAGAGGAAATAAAATTTTAGAAAAAGATTTCACATTATTAATTGCT 76034
```

```
RESULT 13
CR392025
LOCUS       CR392025              158863 bp    DNA     linear   VRT 07-JUN-2005
DEFINITION  Zebrafish DNA sequence from clone CH211-248P17 in linkage group 17,
            complete sequence.
ACCESSION   CR392025
VERSION     CR392025.7  GI:67008317
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 158863)
  AUTHORS   Almeida,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Jun 7, 2005 this sequence version replaced gi:61673560.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            CH211-248P17 is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            Any regions longer than 1kb tagged as misc-feature 'unsure' are
            part of a tandem repeat of more than 10kb in length where it has
            not been possible to anchor the base differences between repeat
            copies.  The region has been built up based on the repeat element
            to match the total size of repeat indicated by restriction digest,
            but repeat copies may not be in the correct order and the usual
            finishing criteria may not apply.
FEATURES             Location/Qualifiers
     source          1. .158863
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="17"
                     /clone="CH211-248P17"
                     /clone_lib="CHORI-211"
ORIGIN

  Query Match             5.7%;  Score 59.4;  DB 5;  Length 158863;
  Best Local Similarity   51.7%;  Pred. No. 1.2;
  Matches  135;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy          4 AATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTA 63
              |||| |  ||  | || |  ||        | |||||     | |   |   ||
Db      89541 AATGATAATAATAATAATAATAATGATAATAATAATGATAATAATAATAATGATAATAAT 89600

Qy         64 CTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTT 123
                 |||  |   ||||| |  || ||  |   || |||||||||  | |    | || |
Db      89601 AATAATAATAATAATAATAATAATAATAATAATAATAATAATGATAATAATAATAATGAT 89660

Qy        124 TGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCA 183
                |||  ||||   |    | |  ||| | |  ||   || ||   ||||  ||||
Db      89661 AATGATAATAATAATAATAATAATGATAATAATAATGATAATAATAAAAATAATAATAAT 89720

Qy        184 AAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAG 243
              || |||||||  |||||  | ||  ||||   ||||| ||||||       || ||||
Db      89721 AATAATAATAATAAAAATAATAATAATGATAATGATGATAATAACAATAATAATAAT 89780

Qy        244 AATTTAAAAAAATGCATTAGT 264
              |||   || ||    | || |
Db      89781 AATAATAATAATAATAATAAT 89801


RESULT 14
AE014822
LOCUS       AE014822              253924 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 14 section 7 of 13 of the
            complete sequence.
ACCESSION   AE014822 AE014187
VERSION     AE014822.1  GI:23497422
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 253924)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 253924)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .253924
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="14"
     repeat_region   68. .123
                     /rpt_type=tandem
     repeat_region   274. .338
                     /rpt_type=tandem
     repeat_region   274. .334
                     /rpt_type=tandem
     repeat_region   442. .471
                     /rpt_type=tandem
     repeat_region   739. .764
                     /rpt_type=tandem
     repeat_region   762. .809
                     /rpt_type=tandem
     repeat_region   1147. .1212
                     /rpt_type=tandem
     repeat_region   1655. .1681
                     /rpt_type=tandem
     repeat_region   2446. .2494
                     /rpt_type=tandem
     repeat_region   2710. .2771
                     /rpt_type=tandem
     repeat_region   3098. .3150
                     /rpt_type=tandem
```

```
repeat_region   3301. .3325
                /rpt_type=tandem
gene            complement(<3843. .>4866)
                /locus_tag="PF14_0353"
mRNA            complement(join(<3843. .4181,4264. .4345,4434. .4627,
                4843. .>4866))
                /locus_tag="PF14_0353"
CDS             complement(join(3843. .4181,4264. .4345,4434. .4627,
                4843. .4866))
                /locus_tag="PF14_0353"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAN36966.1"
                /db_xref="GI:23497423"
                /translation="MMYEHSRQINATLKKCSEMDDMCHLYNMELMQQRERIENLKRKL
                HSYNKNIEEQESIFKNVVKYTDGFINNIITFLERFSNSKELNVLSNCAKYCLADDVDE
                LNLINNFNKVNNISSHKKKEIKTTKLVNNKYKNDYSNKISIYSDENKVNKKIPKKKKT
                KNKYKNDMKKKNVQLNENIYNLHNPYDTVEIRDLYNIYRKKSTTSYENTYIN"
repeat_region   4754. .4817
                /rpt_type=tandem
repeat_region   5303. .5336
                /rpt_type=tandem
repeat_region   5645. .5720
                /rpt_type=tandem
repeat_region   6100. .6133
                /rpt_type=tandem
repeat_region   6611. .6677
                /rpt_type=tandem
repeat_region   6822. .6854
                /rpt_type=tandem
gene            <6975. .>9275
                /locus_tag="PF14_0354"
mRNA            <6975. .>9275
                /locus_tag="PF14_0354"
CDS             6975. .9275
                /locus_tag="PF14_0354"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAN36967.1"
                /db_xref="GI:23497424"
                /translation="MGNTLGIECSFNYVHVTTVLINKKLIKESNNDSKNEKDIREEKE
                KKLPKNVSIPSNDNKLNMNVHLSWLKEKYKKEYINLEEDVSKSDEYNNCDDDYIKMKK
                NTFSYILDMQHIIDSDVQVFSLKKNSEKNVHTHISINELYKCFHEDDNLEKKFIKYLN
                FVKYHKLDISQIDTVNIQHLYDDEIAEIYFWSFKLKYLDECILKYYKKNINYMFINVT
                GKNKNLIKKKFLQITGKNNIFQHNEIKCINNSICFLKRFMPTNLYYFTKYNEENERAS
                CTSQNDEDDEKKKKKKNLLYQSFINKEEVSEIKSYVIVNMKRAVCYHLVNEQNLIERI
                GTLYVGFRTVMGLFLLITGRPCSLQRICQLAKNGTNRTFDMTVQDIYGTSYSNAGLCK
                DLTASFFGNAQHIENVKNILNTYDEEKNINIKEEDDKLMNVYEYENESSCYENVSSCM
                STEISECQEIFETEECIGFEVEKNNINYYRNKYSFLSKDKTKKLLVSNKNFNVYDNCK
                IPIFNFNSDSYCNYMNSFKSKEKLFKHNIEEQQKQEEGYKQKEDDNFLMNDIDNYLSI
                KHSLSDNEINMYEYHRKNYNLKKSFLKKLFKKQYLNENGKIIHNINLSIQKEVSSFIL
                KNKKRFSYKTYDKQIVKKACKSITLKNVCSETNKNKMDSNLKDKTIVKYNTKMCDLSR
                SLLSMAIFTTVYLSYIHCNLYNADHIFFTGYNFEDDVCKELFQIIIKFLSHNRQKIYF
                AKISKYISSLGSAIELINWERIHIDN"
repeat_region   9300. .9332
                /rpt_type=tandem
repeat_region   9569. .9636
                /rpt_type=tandem
gene            <11010. .>11825
                /locus_tag="PF14_0355"
mRNA            <11010. .>11825
                /locus_tag="PF14_0355"
CDS             11010. .11825
                /locus_tag="PF14_0355"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAN36968.1"
                /db_xref="GI:23497425"
                /translation="MPKVNTVKIDNGESDEYNSTNQSPRKLNDSSGLSKKKSGNISRK
                ASSFIYKISKNYKKGVSFVNKKGEYIIGDEDASKITFRDFLNCPLEYHTNVMGTVLFI
                LGSAFFLYPPLYISGCVTFAIACALCLYSNFIGAINKEQRKKNEYYGFIHYIIGCAIF
                ILGSIYSCFENDIYSIITFIIGSLFFLWGSIMFLFSIKFTQIKNVDINILVVFFSNFI
                GGILFTVGSVFFFYPQLYDAACYLFIIGSIIFTLAVYFDYIIYINNNFTLDIV"
repeat_region   11852. .11891
                /rpt_type=tandem
repeat_region   12223. .12248
                /rpt_type=tandem
repeat_region   12422. .12452
                /rpt_type=tandem
repeat_region   12778. .12814
                /rpt_type=tandem
gene            <13191. .>14951
                /locus_tag="PF14_0356"
mRNA            <13191. .>14951
                /locus_tag="PF14_0356"
CDS             13191. .14951
                /locus_tag="PF14_0356"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAN36969.1"
                /db_xref="GI:23497426"
                /translation="MLLYNKRIGKIAIQGRRKCYSWLYFSSYRSNNIIYKEYEDFINI
                GKKNPNQIISMLNRLSKEKRCDIELIKNITNHIYDFNEDYFCPQLIEVLKNYVKLKYS
                DETLLGILSNRIDDLLSIKSCLRTKDMINIYKQLNFHHPVVKDPLLTKFNNNINNYKN
                EMVEIVRNLSYLYIDKDLCEHIFNHIIINYDYYHKDIFSIFEGVSRFDYNDNQTFIQL
                ILKKKIKELEQQNICCNMKDFFKFLSASERLGLNNNTYLHNEFEKKINHIKNISPNNI
                SYILLSMLSTKYRNDTLFELIIMNIENYVNNQNDNNKLLNEPINNKYYNHISSDQYKN
                NESKLNLSNTINGNNNNKYVLNNKRDIITTEQNDYYNVCNSYILNYLPFHLLLLILLN
                YDNKNILLYLLNICINEYTYLYNTSNLIKLLYSNTLLMLPSKIKKYQDKEKVHLEKNV
                LKIFSALQNVYQSCNINDYKILHQCYLFHINMIQNNDTLKHVYKDFLNKECFATLPSS
                YSKLNFEEMQVIRYASMSFLKNKKGHISYYLNKNDFYSSNHNNYDNALLSVKFKVELL
                RKYFDKSEFNVIYNSEKMKI"
repeat_region   15253. .15291
                /rpt_type=tandem
gene            complement(<15469. .>18390)
                /locus_tag="PF14_0357"
mRNA            complement(<15469. .>18390)
                /locus_tag="PF14_0357"
CDS             complement(15469. .18390)
                /locus_tag="PF14_0357"
                /codon_start=1
                /product="acetyl CoA synthetase, putative"
                /protein_id="AAN36970.1"
                /db_xref="GI:23497427"
                /translation="MISKRTNSTFLIIKNYYNVWVGYNKNVQHNSYFKNSLDYIFKPR
                SIGIIGATERKGSVGNSIVDNLIKGESNYKLYFVNSKGSKIYNRQSYKSLKDIKEDSI
                DLAVIAVPRNNVVNVMHELKIKNVRGVVIITAGFKETGAEGLRLENEIINIGKTNNMR
                IIGPNCLGIIHSYHNMNASFASSDILKGHFSLLSQSGAICSAALDLSLQHNIGFSHFI
                SVGSMCDVQFYELVEYLFYDVNTKYILLYVESIGDMNRFVAVCKKVCLYKPIILLKSG
                KTAKAAEAAISHTGSMVGNYEIFYATMKKLGVLVVDNFEELFNMCKVLNSSKYPETNE
                VCVVTNAGGPGVLLVDNIIKNDGVLSNLNDKLKNNLNSFLPDSWSKANPIDILGDASP
                VLYKKTIDAIIKDDQYKNIIVLLSPQSVTEPLDTANEIINLKKQANEQGRLILCNYLG
                GTALEESANMLNKNYIPTFVHPEHSVQNLLKLYENIKHIQGIYEEIPEGLSNSVENHY
                LNKLIEEYYYMKNKSDIHMSPKKKKSIDIIKKAIQNKSYILNEFDSKVILQNYDIPIV
                QTRIFENVDDIKDYDHFAFPCAMKIYSDKITHKKDIGGVILNINNKEELIESYKKIYN
                NVKAHSLEKYFKGVTIQDMVNINDGIEVILGYYYDNNFGPVLLFGSGGSYVEIFKDSV
                LLIPPLNFSYTHHTIKNTKIYKALAGNYSRFTKCDMSKLITTIVKFSELILDLLPYIN
                ECDINPLYVCGDKIIALDARFTLRKNNITDKQENIYYDTFSKIYPIDFISFNDKKKND
                SNFLEGSNNINVSNNIPNETTLNSSLDKIDSITRCIHSYDFKLIHKFCIDNYKELHNS
                TFFYTDVDSIKSELFSYELCNCDYDLYNVILYFNNKQINGLMKIEKRNDNNHCYIYAK
                DNSTFYILIQKLNILAKKQNNKHNYVYLKTNSPYIQDLQSMKYNLEHTYKDVSCYKSN
                N"
repeat_region   18457. .18490
                /rpt_type=tandem
repeat_region   18805. .18890
                /rpt_type=tandem
repeat_region   18985. .19049
                /rpt_type=tandem
repeat_region   19687. .19717
                /rpt_type=tandem
repeat_region   19903. .19938
                /rpt_type=tandem
repeat_region   20047. .20082

Query Match             5.7%;  Score 59.4;  DB 2;  Length 253924;
Best Local Similarity   56.3%;  Pred. No. 1;
```

```
  Matches  111;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy         93 TTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCCACAGAGAGATGA 152
              | | ||  | |||  |||| | |  | || || | |  ||||  || || | | |    |
Db     208926 TATGAATCAAAATAGTTTTTATGAAAATGATTTTAAGAATAATGCTACAAATAAAATGAA 208985

Qy        153 GATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAA 212
              || ||    | |   || ||  ||||    |  |||||||   | |||  | ||  |||
Db     208986 AATGAAGGAAAATGATGATAATCATAATAATATTAATAATATTGATAATAATAATAATAA 209045

Qy        213 TTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCATTAGTTTGACGAC 272
              |   ||| |  |||||      ||| ||||||||  |||||      || |||||| ||
Db     209046 TAATAATAATAATAATCATTATTATTATAAGAATGAAAAAAGTGAAGTTGGTTTGAAGAA 209105

Qy        273 AGGTAAATATTGCATGA 289
              | | || |||   || |
Db     209106 AAGAAAGTATGAAATAA 209122


RESULT 15
BX284625
LOCUS       BX284625               98852 bp    DNA     linear   VRT 02-DEC-2004
DEFINITION  Zebrafish DNA sequence from clone DKEYP-55F12 in linkage group 20
            Contains the gene for a novel protein similar to vertebrate BTB
            (POZ) domain containing family and three novel genes (zgc:55404),
            complete sequence.
ACCESSION   BX284625
VERSION     BX284625.4  GI:36796643
KEYWORDS    HTG; zgc:55404.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 98852)
  AUTHORS   Lovell,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Sep 27, 2003 this sequence version replaced gi:32169085.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            DKEYP-55F12 is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            Any regions longer than 1kb tagged as misc-feature 'unsure' are
            part of a tandem repeat of more than 10kb in length where it has
            not been possible to anchor the base differences between repeat
            copies.  The region has been built up based on the repeat element
            to match the total size of repeat indicated by restriction digest,
            but repeat copies may not be in the correct order and the usual
            finishing criteria may not apply.

FEATURES             Location/Qualifiers
     source          1. .98852
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="20"
                     /clone="DKEYP-55F12"
                     /clone_lib="DanioKeypilot"
     gene            32617. .70238
                     /locus_tag="DKEYP-55F12.1-001"
     mRNA            join(32617. .32698,34133. .35257,56402. .56610,58609. .58684,
                     61088. .61248,62715. .62858,64795. .64984,65541. .65719,
                     67895. .68302,69387. .70238)
                     /locus_tag="DKEYP-55F12.1-001"
                     /product="novel protein similar to vertebrate BTB (POZ)
                     domain containing family"
                     /note="match: ESTs: Em:AI959375.1 Em:BI476036.1
                     Em:BM103068.1"
     CDS             join(32617. .32698,34133. .35257,56402. .56610,58609. .58684,
                     61088. .61248,62715. .62858,64795. .64984,65541. .65719,
                     67895. .68302,69387. .70046)
                     /locus_tag="DKEYP-55F12.1-001"
                     /standard_name="OTTDARP00000005082"
                     /note="match: proteins: Tr:AAH57303 Tr:Q7Z308 Tr:Q80TC3
                     Tr:Q86TS0 Tr:Q8CFE5 Tr:Q9P203"
                     /codon_start=1
                     /product="novel protein similar to vertebrate BTB (POZ)
                     domain containing family"
                     /protein_id="CAH69131.1"
                     /db_xref="GI:55251035"
                     /translation="MGVNASSYPHSCSPRFGGNPQTQQTFIGASYAPQGYGGESKLYS
                     LEHGPEKPQDKKKKSSGLATLKRRFIKRRKSSRSADHARQMRELLSGWDVRDVNALVE
                     EYEGTAALKELSLQAGLARPEARTLQRDLATLYQHKYCTDVDLIFQDSCFPAHRAILA
                     TRCPFFKTLLSSSPGYGAEVLLDVGTAGMDAPMFSSLLHYLYTGELGSDDARLQNVDV
                     LVRLSEEFGTPNSLEADMRNLCEHMPYFDSLLSFSSDSELVEAFGAGGPAPSSMGSGN
                     IGASHGSPDEELRAHKAVLSARSPFFRNLLQRRIRSGEEITERTLHTPTRIILDESII
                     PKKYARVVLHCMYTDAVDLSLVLRGSPSEGSLGEVQALVAGRGCMSRAEEAMELYHIA
                     LFLEFSMLAQGCEDIVADSISLDSVVAILKWSSQPYGSKWVHRQAMHFLCEEFSQVMT
                     SDVLYELSKEHLLAAIQSDYLQASEQDILKYVVKWGEHQLIKRMADREPNLLSGTAHS
                     VNKRGVKRRDLDLEELKEILSPLLPFVRTEHILPANSEVLSDTMKRGLISTPPSDMLP
                     TAEGGKANAWLRQKSAGIYVRPRLFSPYVEEAKAVLEEMMVEQTDLVRLRMVRMSNVP
                     DTLYMVNNAVPQCCHMINHQQMAGSQSAPPSVVANEIPVPRLAVVKEMIRRLQELRHT
                     DQVQRAFALNCGEGATVSYELQIRVLREFGLPDGAAELLQNPHKFFPEERFGDESPVL
                     ALRQSGRCRVNSTPAVESMFTDLEALVGFHPPLPPPPPPPYHPPTTPGHGQQQFKAGWR
                     PRVPSQPPSRSFSYPCNHALLHRQASSKMGSPVYPPGAKPLPPDCTNAQGGRNLHPDK
                     QMNMEPVINEFMPDIAMGVSAMTLKERRLPEVSMEVESHEPSHPHLSTAPPMSHSYSN
                     RHPHSSRKRHAAEPKPEGPQPDFPDLYEFPGRHVPPYSGPDLYIHGRQAGGGAPPPSY
                     SEASESLRLDVLEQPPQRLDLAIVSQGSVVQAIQRPRPINETDLTCGLGPSTEQYEEW
                     PTGRRPPPEGIGLGLGANVGTGDEGGLGARDRRSPNKPEYPYRKSAL"
     polyA_site      complement(75403)
     polyA_signal    complement(75424. .75429)
     gene            complement(77403. .92070)
                     /locus_tag="DKEYP-55F12.2-001"
     mRNA            complement(join(77403. .78151,80523. .80584,80683. .80863,
                     82664. .82855,85044. .85165,88630. .88735,88850. .88903,
                     89005. .89100,89185. .89245,90381. .90514,91705. .92070))
                     /locus_tag="DKEYP-55F12.2-001"
                     /product="novel protein (zgc:55404)"
                     /note="match: cDNAs: Em:BC044390.1
                     match: ESTs: Em:AI545120.1 Em:BM104712.1"
     CDS             complement(join(78056. .78151,80523. .80584,80683. .80863,
                     82664. .82855,85044. .85165,88630. .88735,88850. .88903,
                     89005. .89100,89185. .89245,90381. .90514,91705. .91833))
                     /locus_tag="DKEYP-55F12.2-001"
                     /standard_name="OTTDARP00000005066"
                     /note="match: proteins: Tr:Q803P8"
                     /codon_start=1
                     /product="novel protein (zgc:55404)"
                     /protein_id="CAH69132.1"
                     /db_xref="GI:55251036"
                     /translation="MAANDGEEATVSLVDVLEEDEELENEASAVLGGSDSEKCSYPEG
                     YVKRQALYACNTCTPKGGEPAGICLACSYKCHEGHDLFELYTKRNFRCDCGNDKFGDM
                     ECKLFADKEKVNSGNKYSHNFFGLYCTCDRPYPDPEDEVSDEMIQCIVCEDWLHGRHL
```

```
                     GCAVPDCVELQEMICESCMNKTPFLWNYAAQIAVSPIIKVSPCKAEEDVKVEVEKEEE
                     EKMVKNENTKSAVDEKVQMNGTSGCKRKLQDESEGSSGHGLCSGCRLREMKASGISRK
                     QQGAVFWPSAWRTKLCSCPDCKVLYADAGLCFLTDEADTVLAYENKGKSTEQAEQAGG
                     AGGRDPLMSALNNLNRVQQLEIIHEYNDMKTELKDYLQRFAAEGKVVTPEDIRQFFEQ
                     QQSRKRRRANAGQPYYCS"
     gene            92822. .94964
                     /locus_tag="DKEYP-55F12.3-001"
     mRNA            join(92822. .93029,93802. .94964)
                     /locus_tag="DKEYP-55F12.3-001"
                     /product="novel protein"
                     /note="match: ESTs: Em:AI959455.1 Em:AL923891.1
                     Em:BG303822.1 Em:CK399222.1"
     CDS             join(92834. .93029,93802. .93914)
                     /locus_tag="DKEYP-55F12.3-001"
                     /standard_name="OTTDARP00000005070"
                     /codon_start=1
                     /product="novel protein"
                     /protein_id="CAH69133.1"
                     /db_xref="GI:55251037"
                     /translation="MAACELKGELKYRDGQTNQFTAQVDGNLKSIIGGIKKINGDISE
                     ALTTLVEQEKGSTGNDRESVNDDEEEDSDDDGSATKTTTITKSSSEPPSKKMKTLRS
                     "
     misc_feature    94853
                     /locus_tag="DKEYP-55F12.3-001"
                     /note="Clone_left_end: DKEY-177P2"
     gene            complement(95050. .97573)
                     /locus_tag="DKEYP-55F12.4-001"
     mRNA            complement(join(95050. .96103,97470. .97573))
                     /locus_tag="DKEYP-55F12.4-001"
                     /product="novel protein"
                     /note="match: ESTs: Em:BQ078747.1 Em:CB366378.1
                     Em:CK026964.1 Em:CK028391.1"
     CDS             complement(95660. .96088)
                     /locus_tag="DKEYP-55F12.4-001"
                     /standard_name="OTTDARP00000005074"
                     /note="match: proteins: AAH30119 Tr:Q8BH26 Tr:Q8N6I4
                     Tr:Q9Y4S5"
                     /codon_start=1
                     /product="novel protein"
                     /protein_id="CAH69134.1"
                     /db_xref="GI:55251038"
                     /translation="MMNFRQRMGWIGVGLYLLASVAAVYYIFEISQTYNRLALAQVEK
                     TSGAQAKWPGDASSSSPSSTSWIVTLKTRLLLLPFWVWATIFLLPYLQVFLFLYSCTR
                     ADPKTVGYCILPICLAVLCNRHQTFTKASNQISRLQLIDT"
ORIGIN

  Query Match             5.7%;  Score 59.2;  DB 5;  Length 98852;
  Best Local Similarity   52.0%;  Pred. No. 1.5;
  Matches  133;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy          1 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 60
              || | | || |||| | ||||  |||       | ||||||    | |   |   |
Db      32076 ATTATTATTATTATTATTATTATTATTATTAATAATAATAATAATAATAATAATAATAAT 32135

Qy         61 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 120
                 ||| ||   ||||| |  || ||||| | || | || |  | |||  || |||
Db      32136 AATAATAATAACAATAATAATAATAATAATTGTATTAATATTATTATTGTTGTTGTTATT 32195

Qy        121 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 180
              |   |   |   ||    | |   |||| |  |  | |  ||  |  || |  | ||
Db      32196 CCTATAGCTAAGGTATAAAAATATGCAGATAAATTTTAATGGATCCTTAAATATAAAAAC 32255

Qy        181 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 240
              |  || || | ||  || |||||  | |||   || |  || ||  |   |||| |
Db      32256 TATTATTATTATCATCATTATTTTTATTATGATTATTATTATCATTATAGATTTTATGGT 32315

Qy        241 AAGAATTTAAAAAAAT 256
              || |||||||| |||
Db      32316 AATAATTTAAAATAAT 32331


Search completed: May 11, 2006, 19:14:30
Job time : 3762.83 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 07:09:59 ; Search time 475.701 Seconds
(without alignments)
14626.705 Million cell updates/sec

Title: US-09-423-093B-3_COPY_3977_5020
Perfect score: 1044
Sequence: 1 atgaatgtttttatcagtat.....ctccgcgtcttggtatttaa 1044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1044	100.0	12441	2	AAX06750	Aax06750 Salmonell
	2	56.2	5.4	637	1	AAN90220	Aan90220 Malaria-s
	3	56.2	5.4	5139	3	AAA70139	Aaa70139 Plasmodiu
	4	55.8	5.3	1404	8	ACA28205	Aca28205 Prokaryot
c	5	53.6	5.1	6642	4	ABL05270	Abl05270 Drosophil
c	6	53.6	5.1	10992	4	ABL05242	Abl05242 Drosophil
c	7	52.8	5.1	4985	6	ABQ75107	Abq75107 Anopheles
c	8	52.8	5.1	4985	10	ACF79720	Acf79720 Mosquito
	9	51.8	5.0	2000	11	ACL37108	Acl37108 Rice stre
	10	51.8	5.0	5940	3	AAA70105	Aaa70105 Plasmodiu
	11	51.6	4.9	11143	4	ABL12834	Abl12834 Drosophil
	12	51	4.9	7722	3	AAA70168	Aaa70168 Plasmodiu
	13	50.8	4.9	337344	13	ABD32715	Abd32715 Human can
	14	50.6	4.8	3567	3	AAA70117	Aaa70117 Plasmodiu
c	15	50.4	4.8	2000	11	ACL35887	Acl35887 Rice stre
	16	49.8	4.8	3501	3	AAA70202	Aaa70202 Plasmodiu
c	17	49.6	4.8	2000	11	ACL37108	Acl37108 Rice stre
	18	49.4	4.7	11922	3	AAA70187	Aaa70187 Plasmodiu
	19	49.2	4.7	999	3	AAA70234	Aaa70234 Plasmodiu
	20	49	4.7	6033	3	AAA70152	Aaa70152 Plasmodiu
	21	48.8	4.7	2000	11	ACL35363	Acl35363 Rice stre
	22	48.8	4.7	3927	3	AAA70101	Aaa70101 Plasmodiu
c	23	48.8	4.7	6179	4	AAS46344	Aas46344 Tumour su
c	24	48.8	4.7	6179	6	ABK31251	Abk31251 Signal tr
	25	48.2	4.6	3315	14	ADW71859	Adw71859 OKT3Vk an
	26	48.2	4.6	8801	4	AAS45437	Aas45437 Chemicall
	27	48.2	4.6	8801	6	ABL33741	Abl33741 Human imm
	28	48.2	4.6	8801	6	ABK28286	Abk28286 DNA trans
c	29	48.2	4.6	9963	6	ABL32695	Abl32695 Human imm
	30	48	4.6	1340	14	ADZ71035	Adz71035 Human chr
	31	48	4.6	50000	6	ABL55644	Abl55644 AmEPV gen
	32	47.8	4.6	651	14	ADZ71113	Adz71113 Human chr
	33	47.8	4.6	1344	3	AAA70191	Aaa70191 Plasmodiu
c	34	47.6	4.6	2000	11	ACL35363	Acl35363 Rice stre
	35	47.6	4.6	5376	6	ABL34150	Abl34150 Human imm
c	36	47.6	4.6	11812	4	AAS45501	Aas45501 Chemicall
c	37	47.6	4.6	11812	4	AAS46741	Aas46741 Tumour su
c	38	47.6	4.6	11812	6	ABL34118	Abl34118 Human imm
c	39	47.6	4.6	11812	6	ABK28431	Abk28431 DNA trans
	40	47.4	4.5	7108	6	ABK39995	Abk39995 Human che
	41	47.4	4.5	18624	6	ABL33702	Abl33702 Human imm
c	42	47.2	4.5	5340	5	AAF26289	Aaf26289 P. falcip
c	43	47.2	4.5	7508	6	ABK31207	Abk31207 Signal tr
	44	47.2	4.5	32392	6	ABL56203	Abl56203 AmEPV gen
c	45	47.2	4.5	50000	6	ABL55643	Abl55643 AmEPV gen

ALIGNMENTS

RESULT 1
AAX06750
ID AAX06750 standard; DNA; 12441 BP.
XX
AC AAX06750;
XX
DT 26-APR-1999 (first entry)
XX
DE Salmonella enterica serovar C2 O antigen gene cluster.
XX
KW O antigen; diagnosis; wbaR gene; wzx gene; wbaL gene; wbaQ gene;
KW wxy gene; wbaW gene; wbaZ gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Salmonella enterica.
XX
FH Key Location/Qualifiers
FT CDS 138. .1019
FT /*tag= a
FT /note= "abe gene; encodes AAW88324"
FT CDS 1019. .2359
FT /*tag= b
FT /product= "flippase"
FT /note= "wzx gene, this region is specifically claimed in
FT Claim 9; encodes AAW88325"
FT CDS 2362. .3314
FT /*tag= c
FT /product= "abequosyl_transferase"
FT /note= "wbaR gene, this region is specifically claimed in
FT Claim 9; encodes AAW88326"
FT CDS 3361. .3975
FT /*tag= d
FT /product= "acetyl_transferase"
FT /note= "wbaL gene, this region is specifically claimed in
FT Claim 9; encodes AAW88327"
FT CDS 3977. .5020
FT /*tag= e
FT /product= "rhamnosyl_transferase"
FT /note= "wbaQ gene, this region is specifically claimed in
FT Claim 9; encodes AAW88328"
FT CDS 5114. .6313
FT /*tag= f

```
FT                   /product= "O_antigen_polymerase"
FT                   /note= "wzy gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88329"
FT   CDS             6313. .7323
FT                   /*tag= g
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88330"
FT   CDS             7310. .8467
FT                   /*tag= h
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88331"
FT   CDS             8482. .9903
FT                   /*tag= i
FT                   /note= "manC gene; encodes AAW88332"
FT   CDS             9911. .11347
FT                   /*tag= j
FT                   /note= "manB gene; encodes AAW88333"
FT   CDS             11426. .12441
FT                   /*tag= k
FT                   /note= "wbaP gene; encodes AAW88334"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88324, AAW88325, AAW88326, AAW88327, AAW88328, AAW88329,
DR   AAW88330, AAW88331, AAW88332, AAW88333, AAW88334.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 9; Fig 9; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Salmonella enterica serotype C2 (strain M67) O
CC   antigen. 11 Open reading frames within the gene cluster were identified
CC   on the basis of homology to known sequences. The genes encode proteins
CC   (see AAW88323-34) involved the synthesis of sugars present in the
CC   polysaccharide antigen, and in the transport or processing of
CC   polysaccharide or oligosaccharide units. The use of nucleic acid
CC   molecules derived from assembly and transport genes, particularly wbd
CC   (transferase), wzx (flippase) and wzy (polymerase) genes, within O
CC   antigen gene clusters improves the specificity of methods for the
CC   detection and identification of O antigens, e.g. in testing food- or
CC   faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of S. enterica
XX
SQ   Sequence 12441 BP; 3887 A; 1876 C; 2518 G; 4160 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1044;  DB 2;  Length 12441;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-210;
  Matches 1044;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3977 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 4036

Qy         61 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4037 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 4096

Qy        121 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4097 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 4156

Qy        181 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4157 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 4216

Qy        241 AAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4217 AAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAAC 4276

Qy        301 GATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4277 GATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCT 4336

Qy        361 GAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4337 GAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGT 4396

Qy        421 GATACTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4397 GATACTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAA 4456

Qy        481 TTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4457 TTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAA 4516

Qy        541 AAACTATCGAGTGATTTATTTGATGGGCGTTTATATTATCAAATGTACCTTGCTGGTATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4517 AAACTATCGAGTGATTTATTTGATGGGCGTTTATATTATCAAATGTACCTTGCTGGTATG 4576

Qy        601 CTAATGGCTGAAGGTCAGGGATACTATTTTAGCGACGTGATGACATTGTCGAGGGATACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4577 CTAATGGCTGAAGGTCAGGGATACTATTTTAGCGACGTGATGACATTGTCGAGGGATACA 4636

Qy        661 GAGGCTCCTGACTTTGGTAACGCTGGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4637 GAGGCTCCTGACTTTGGTAACGCTGGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTAT 4696

Qy        721 AAACCAGAGGGCCGTATACATATGGTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4697 AAACCAGAGGGCCGTATACATATGGTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAA 4756

Qy        781 GATACAACAAAAATTGATGGCGTTTATGCTGGAATTAGAAAAGACTTAGCGAACTATTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4757 GATACAACAAAAATTGATGGCGTTTATGCTGGAATTAGAAAAGACTTAGCGAACTATTTT 4816

Qy        841 TATCCTTATATTCGAGATCAACTCGACTTGCCTCTTTATACTTATATTAAAATGATAAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4817 TATCCTTATATTCGAGATCAACTCGACTTGCCTCTTTATACTTATATTAAAATGATAAAT 4876

Qy        901 AAATTTCGGAAAATGGGATTTTCAAATGAAAAGCTTTTCTATGTGCATGCCTTTTTAGGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4877 AAATTTCGGAAAATGGGATTTTCAAATGAAAAGCTTTTCTATGTGCATGCCTTTTTAGGG 4936

Qy        961 TATGTACTAAAACGGAGGGGCTATGATGCTTTAATTAAATACATTCGTAGCAAAAAAGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4937 TATGTACTAAAACGGAGGGGCTATGATGCTTTAATTAAATACATTCGTAGCAAAAAAGGC 4996

Qy       1021 GGTACTCCGCGTCTTGGTATTTAA 1044
              ||||||||||||||||||||||||
Db       4997 GGTACTCCGCGTCTTGGTATTTAA 5020


RESULT 2
AAN90220
ID   AAN90220 standard; DNA; 637 BP.
XX
AC   AAN90220;
XX
DT   25-MAR-2003  (revised)
```

DT 01-NOV-1989 (first entry)
XX
DE Malaria-specific DNA insert of clone 41-13.
XX
KW DNA; malaria; clone 41-13; Plasmodium falciparum; protein; vaccine.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 3. .635
FT /*tag= a
XX
PN EP322712-A.
XX
PD 05-JUL-1989.
XX
PF 20-DEC-1988; 88EP-00121299.
XX
PR 30-DEC-1987; 87DE-03744495.
PR 15-SEP-1988; 88DE-03831351.
XX
PA (BEHW) BEHRINGWERKE AG.
XX
PI Knapp B, Hundt E, Enders B, Kupper H;
XX
DR WPI; 1989-194071/27.
DR P-PSDB; AAP90412.
XX
PT New antigenic proteins from Plasmodium falciparum - new encoding nucleic
PT acid sequences and derived antibodies, useful in vaccines, diagnosis,
PT etc.
XX
PS Claim 1; Page 23 and Table 12; 25pp; German.
XX
CC The DNA is of a malaria-specific DNA insert of clone 41-13. The encoded
CC peptide (see AAP90412) is useful in a protective vaccine, esp. against
CC malaria. Produced antibodies are useful for passive immunisation, and
CC antibodies, the protein and the DNA sequence are useful as diagnostic
CC reagents. See also AAN90211-25 and AAP90403-19. (Updated on 25-MAR-2003
CC to correct PR field.) (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 637 BP; 291 A; 52 C; 106 G; 188 T; 0 U; 0 Other;

Query Match 5.4%; Score 56.2; DB 1; Length 637;
Best Local Similarity 49.8%; Pred. No. 0.03;
Matches 142; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Qy 4 AATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTA 63
Db 78 AATGATCTTTTAAATGTTTATATATCACCTAATATGATTAATCATTCTTTATCTTCAACT 137

Qy 64 CTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTT 123
Db 138 TGTGAAAAAAAAAATAAAGAAGATAACAAAATGAATGACAATAAATTTCTTAATAGCAGT 197

Qy 124 TGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCA 183
Db 198 AGTAAAATGAAAATTCCAGAGATAAGTACGAACAACTCAAATGAAAAGATTGTTAATGTG 257

Qy 184 AAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAG 243
Db 258 TCAAATGATGAAATGTTAGTATATCATAATTTAACCGTATTAAATGTAAAGGAACAAGGA 317

Qy 244 AATTTAAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATG 288
Db 318 GGTGTAACAGAAGAATCGAGCTGTATAAAACGCACATATTTTGTG 362

RESULT 3
AAA70139
ID AAA70139 standard; DNA; 5139 BP.
XX
AC AAA70139;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:272.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US026796.
XX
PR 05-NOV-1998; 98US-0107131P.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection.
XX
PS Disclosure; Page 482-484; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
CC (II) are useful for the development of vaccines against P. falciparum
CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
CC immunogens comprising the sequences of (I), are useful in the detection
CC of infection with P. falciparum. Furthermore, (I) (especially when they
CC are rifins or secreted or membrane proteins) can aid the identification
CC of drugs to treat or prevent P. falciparum infection, or they can be used
CC to identify drug resistance in P. falciparum. Sequencing of the
CC Plasmodium chromosome 2 and the subsequent identification of proteins
CC encoded by it will help to expand our understanding of parasite biology,
CC a process hampered by the complexity of the parasitic lifecycle, and
CC provide new targets for vaccine and drug development. Parasite resistance
CC to drugs and mosquito resistance to insecticides have led to a resurgence
CC of malaria in many parts of the world, and there is a pressing need for
CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
CC represent nucleotide and protein sequences given in the present
CC invention, but which are not specifically mentioned within the
CC specification
XX
SQ Sequence 5139 BP; 2440 A; 434 C; 712 G; 1553 T; 0 U; 0 Other;

Query Match 5.4%; Score 56.2; DB 3; Length 5139;
Best Local Similarity 49.8%; Pred. No. 0.043;
Matches 142; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Qy 4 AATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTA 63
Db 1978 AATGATCTTTTAAATGTTTATATATCACCTAATATGATTAATCATTCTTTATCTTCAACT 2037

Qy 64 CTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTT 123
Db 2038 TGTGAAAAAAAAAATAAAGAAGATAACAAAATGAATGACAATAAATTTCTTAATAGCAGT 2097

Qy 124 TGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCA 183
Db 2098 AGTAAAATGAAAATTCCAGAGATAAGTACGAACAACTCAAATGAAAAGATTGTTAATGTG 2157

```
Qy        184 AAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAG 243
              |||| || ||       | |||  |||||| |  | |   ||| ||    |  |
Db       2158 TCAAATGATGAAATGTTAGTATATCATAATTTAACCGTATTAAATGTAAAGGAACAAGGA 2217

Qy        244 AATTTAAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATG 288
              | ||| | ||      || |  |  | | | | |||||   ||
Db       2218 GGTGTAACAGAAGAATCGAGCTGTATAAAACGCACATATTTTGTG 2262


RESULT 4
ACA28205
ID   ACA28205 standard; DNA; 1404 BP.
XX
AC   ACA28205;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Prokaryotic essential gene #9862.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Clostridium botulinum.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU24335.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 16075; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC   prokaryotic essential genes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1404 BP; 633 A; 103 C; 236 G; 432 T; 0 U; 0 Other;

  Query Match             5.3%;  Score 55.8;  DB 8;  Length 1404;
  Best Local Similarity   54.0%;  Pred. No. 0.042;
  Matches  114;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy         74 TATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATA 133
              |||||||  ||||| ||| | |||||||      ||  ||         ||    | ||
Db        965 TATATAAAGAAGATAATTTTATAAAGAAGGTAAATAATGTTCAGTGTGAACTTAAAATAG 1024

Qy        134 AATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATA 193
               || |   ||   ||| ||  ||||   || | |  |||  |||     |||| || |||
Db       1025 GATTTAATCAAGTAGAAGATTTAAATATTAATTTAAAAATATATTCTATAAAAGATCATA 1084

Qy        194 AACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAA 253
              |  | ||| |  ||| ||||    ||  | |  |||| |  ||||||||      ||
Db       1085 ATTATAATATAAATGATAATAAATATATACAGTATTTTGATTATGATAAAGTAAGTAATG 1144

Qy        254 AATGCATTAGTTTGACGACAGGTAAATATTG 284
              ||   |||  ||| |  |   | ||| || |
Db       1145 AAAAAATTTATTTAAGAAATAGAAAAAATGG 1175


RESULT 5
ABL05270/c
ID   ABL05270 standard; cDNA; 6642 BP.
XX
AC   ABL05270;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 10292.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB61167.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 10292; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
```

CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6642 BP; 1890 A; 1542 C; 1546 G; 1664 T; 0 U; 0 Other;

Query Match 5.1%; Score 53.6; DB 4; Length 6642;
Best Local Similarity 55.3%; Pred. No. 0.16;
Matches 104; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              ||| ||   ||||| |  || ||  |   || |||||| ||  | |    | |  |  |
Db       6632 ATACCAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 6573

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |  || | |  ||   || ||   | ||  ||||    || |
Db       6572 ATAATAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAACAATA 6513

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| ||  |||||      || |||||||
Db       6512 ATAATAATAATAATAATAATAATAATAATAATAAAATTAATTATAATGATAATAAGAACT 6453

Qy        248 TAAAAAAA 255
              ||| || |
Db       6452 TAACAACA 6445
```

RESULT 6
ABL05242/c
ID ABL05242 standard; cDNA; 10992 BP.
XX
AC ABL05242;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10208.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB61139.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 10208; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 10992 BP; 3244 A; 2451 C; 2349 G; 2948 T; 0 U; 0 Other;

Query Match 5.1%; Score 53.6; DB 4; Length 10992;
Best Local Similarity 55.3%; Pred. No. 0.17;
Matches 104; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              ||| ||   ||||| |  || ||  |   || |||||| ||  | |    | |  |  |
Db      10982 ATACCAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 10923

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |  || | |  ||   || ||   | ||  ||||    || |
Db      10922 ATAATAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAACAATA 10863

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| ||  |||||      || |||||||
Db      10862 ATAATAATAATAATAATAATAATAATAATAATAAAATTAATTATAATGATAATAAGAACT 10803

Qy        248 TAAAAAAA 255
              ||| || |
Db      10802 TAACAACA 10795
```

RESULT 7
ABQ75107/c
ID ABQ75107 standard; cDNA; 4985 BP.
XX
AC ABQ75107;
XX
DT 01-NOV-2002 (first entry)
XX
DE Anopheles gambiae odourant receptor 2 genomic DNA SEQ ID NO:10.
XX
KW Anopheles gambiae; mosquito; olfactory gene; arrestin 1; pest control;
KW odourant receptor; olfaction; gene; ds.
XX
OS Anopheles gambiae.
XX
PN WO200259274-A2.
XX
PD 01-AUG-2002.
XX
PF 28-JAN-2002; 2002WO-US002549.
XX
PR 26-JAN-2001; 2001US-0264649P.
PR 24-JAN-2002; 2002US-00056405.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Zwiebel LJ;
XX
DR WPI; 2002-627421/67.
DR P-PSDB; ABP52835.
XX
PT New mosquito olfaction polypeptides and polynucleotides, useful for
PT mosquito management, i.e. controlling the pest and disease vectors, or
PT for identifying pest control agents.
XX
PS Disclosure; Fig 4a; 96pp; English.
XX
CC The present invention describes a purified Anopheles gambiae olfaction
CC polypeptide comprising a 383, 394, 380, 411, 412, 391, 157 or 401 residue
CC amino acid sequence (see ABP52833 to ABP52840) (S1), a conservatively
CC modified amino acid sequence of them, or a sequence of (S1) with at least
CC 20 consecutive residues. Also described: (1) an isolated polynucleotide
CC comprising: (a) a nucleotide sequence encoding the purified Anopheles
CC gambiae olfaction polypeptide; or (b) a nucleotide sequence that
CC hybridises under stringent conditions to a hybridisation probe comprising

```
CC   a 1964, 1239, 1142, 1236, 1194, 1176, 474 or 1206 nucleotide sequence
CC   (see ABQ75102 to ABQ75105 and ABQ75110 to ABQ75113) (S2), or its
CC   complement; and (2) a method for identifying an agent that binds to
CC   mosquito olfaction molecules comprising: (a) providing an isolated
CC   mosquito olfaction molecule; (b) contacting a test agent with the
CC   isolated mosquito olfaction molecule; and (c) detecting specific binding
CC   of the test agent to the isolated mosquito olfaction molecule, where the
CC   presence of specific binding identifies the test agent as a mosquito
CC   olfaction-binding compound. The mosquito olfaction molecules are useful
CC   for mosquito management, i.e. controlling this pest and disease vector. A
CC   method from the present invention of screening for substances that
CC   modulate arrestin-odourant receptor interaction is useful for identifying
CC   pest control agents. The present sequence represents Anopheles gambiae
CC   odourant receptor 2 genomic DNA from the present invention. N.B. The
CC   features given in figure 4a are tentative and do not directly encode SEQ
CC   ID NO:6 (ABP52835)
XX
SQ   Sequence 4985 BP; 1319 A; 1058 C; 990 G; 1618 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 52.8;  DB 6;  Length 4985;
  Best Local Similarity   54.0%;  Pred. No. 0.22;
  Matches  108;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||  |   |||||  |  || ||  |   || ||||||  ||  | |    | | ||  |
Db       2535 ATAAAAGTAATAATTATAATAATAATAGTAATAATAATAATAATAATAATAATTAATA 2476

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |   || | |  ||   || |    | ||  ||||    || |
Db       2475 ATAATAATAATAATAATAATAATAATAATAATAATAATTATAATAATAATAATAATAATA 2416

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| |  |||||       || |||| |||
Db       2415 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2356

Qy        248 TAAAAAAATGCATTAGTTTG 267
                 || |||| |  || ||||
Db       2355 ATAATAAATTCTGTATTTTG 2336


RESULT 8
ACF79720/c
ID   ACF79720 standard; DNA; 4985 BP.
XX
AC   ACF79720;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Mosquito odorant receptor 2 genomic DNA.
XX
KW   Mosquito; odorant receptor 2; olfaction; insecticide; antimalarial; gene;
KW   ss.
XX
OS   Anopheles gambiae.
XX
FH   Key             Location/Qualifiers
FT   exon            1. .702
FT                   /*tag= d
FT                   /number= 1
FT   5'UTR           1. .485
FT                   /*tag= a
FT   TATA_signal     295. .300
FT                   /*tag= b
FT   CDS             486. .3561
FT                   /*tag= c
FT                   /product= "Odorant receptor 2"
FT                   /note= "contains introns"
FT   intron          703. .798
FT                   /*tag= e
FT                   /number= 1
FT   exon            799. .882
FT                   /*tag= f
FT                   /number= 2
FT   intron          883. .978
FT                   /*tag= g
FT                   /number= 2
FT   exon            979. .1404
FT                   /*tag= h
FT                   /number= 3
FT   intron          1405. .1601
FT                   /*tag= i
FT                   /number= 3
FT   exon            1602. .1702
FT                   /*tag= j
FT                   /number= 4
FT   intron          1703. .1783
FT                   /*tag= k
FT                   /number= 4
FT   exon            1784. .1886
FT                   /*tag= l
FT                   /number= 5
FT   intron          1887. .1993
FT                   /*tag= m
FT                   /number= 5
FT   exon            1994. .2106
FT                   /*tag= n
FT                   /number= 6
FT   intron          2107. .3459
FT                   /*tag= o
FT                   /number= 6
FT   exon            3460. .4985
FT                   /*tag= p
FT   3'UTR           3562. .4985
FT                   /*tag= q
XX
PN   WO2003076590-A2.
XX
PD   18-SEP-2003.
XX
PF   10-MAR-2003; 2003WO-US007174.
XX
PR   08-MAR-2002; 2002US-00094240.
XX
PA   (UYVA-) UNIV VANDERBILT.
XX
PI   Zwiebel LJ;
XX
DR   WPI; 2003-722331/68.
XX
PT   New mosquito arrestin 1 and 2 genes and polypeptides, useful for
PT   identifying mosquito olfaction molecule binding compounds which reduce
PT   the ability of mosquitoes to locate sources of bloodmeal, e.g. humans.
XX
PS   Disclosure; Fig 4a; 101pp; English.
XX
CC   The present sequence is that of genomic DNA encoding a novel mosquito
CC   olfaction molecule, odorant receptor 2. A cDNA sequence for odorant
CC   receptor 2 is given in ACF79719. The invention provides 9 novel mosquito
CC   polypeptides and the nucleic acids encoding them. These are odorant
CC   receptor molecules 1-7 and arrestins 1 and 2. The odorant receptors
CC   function in a ligand-induced signal transduction pathway for the
CC   activation of mosquito olfaction. Arrestin functions to inhibit the
CC   activated signal transduction cascade. Thus, the odorant receptors act as
CC   an 'on' switch, and arrestin as an 'off' switch for the odorant detection
CC   system of the mosquito. Methods are provided for identifying compounds
CC   that interfere with the operation of the mosquito olfactory system,
CC   particularly compounds that modulate arrestin 2 activity. These are
CC   useful for the control of mosquitoes, particularly by reducing their
CC   ability to locate sources of bloodmeal
XX
SQ   Sequence 4985 BP; 1319 A; 1058 C; 990 G; 1618 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 52.8;  DB 10;  Length 4985;
  Best Local Similarity   54.0%;  Pred. No. 0.22;
  Matches  108;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;
```

```
Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||  |   |||||  |  || ||  |   || |||||| ||  | |    | | ||  |
Db       2535 ATAAAAGTAATAATTATAATAATAATAGTAATAATAATAATAATAATAATAATTAATA 2476

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |   || | |  ||   || |    | ||  ||||    || |
Db       2475 ATAATAATAATAATAATAATAATAATAATAATAATAATTATAATAATAATAATAATAATA 2416

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| |  ||||       || |||| |||
Db       2415 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2356

Qy        248 TAAAAAAATGCATTAGTTTG 267
                 || |||| |  || ||||
Db       2355 ATAATAAATTCTGTATTTTG 2336


RESULT 9
ACL37108
ID   ACL37108 standard; cDNA; 2000 BP.
XX
AC   ACL37108;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Rice stress-regulated promoter SEQ ID NO:15671.
XX
KW   ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW   agriculture.
XX
OS   Oryza sativa.
XX
PN   WO2003008540-A2.
XX
PD   30-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US019668.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   24-AUG-2001; 2001US-0314662P.
PR   26-SEP-2001; 2001US-0325277P.
PR   21-NOV-2001; 2001US-0332132P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Kreps J,  Briggs SP,  Cooper B,  Glazebrook J,  Goff SA,  Katagiri F;
PI   Moughamer T,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-248011/24.
XX
PT   New stress-responsive nucleic acid, useful for altering the
PT   responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT   stress, salt stress or osmotic stress.
XX
PS   Claim 48; SEQ ID NO 15671; 89pp; English.
XX
CC   The invention relates to novel abiotic stress responsive polynucleotides
CC   and polypeptides. Also disclosed are vectors, expression cassettes, host
CC   cells, and plants containing such polynucleotides. Also disclosed are
CC   methods for using the polynucleotides and polypeptides to alter the
CC   responsiveness of a plant to abiotic stress. The invention is useful in
CC   agriculture. The nucleic acid is useful for determining whether a test
CC   plant has been exposed to an abiotic stress condition. It is also useful
CC   for selecting an agent that alters abiotic stress regulated
CC   polynucleotide expression in a plant cell, and to identify a homolog or
CC   ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC   molecule and the polypeptide encoded by it are useful in altering the
CC   responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC   stress, osmotic stress or any of their combinations. The present sequence
CC   is used in the exemplification of the invention
XX
SQ   Sequence 2000 BP; 510 A; 350 C; 268 G; 518 T; 0 U; 354 Other;

  Query Match             5.0%;  Score 51.8;  DB 11;  Length 2000;
  Best Local Similarity   12.7%;  Pred. No. 0.31;
  Matches   49;  Conservative  180;  Mismatches  157;  Indels    1;  Gaps    1;

Qy        171 AAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTT 230
              ::: ::: :: :: |: :: |:  ::: :::: |  :     ::: :: : ::: || ::
Db         67 MMRWYWKKMWAMWYAWCMWYAYGYMMWKWWYSRTMTYTRAYYKKWKMWRRKMSRMAAAWY 126

Qy        231 AGGCTATGATAAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGAT 290
              ::::|::: :    ::::::|::: ::|:: :|:::::::  ::: :: :: :   |
Db        127 YWSYWRTSRWKWTGYRKKRWWRARRYMYWTYWCYTWSWMRWYRYWKMMKYRAYKWMAYTT 186

Qy        291 CATGGGCAACGATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAA 350
                : |::   |  : : : |  |:|:: ::    :    :: :| |  :::   |  :
Db        187 TYMTKCCMWWWCTYGYKWKWAAWRASAKRARRAWMMMRWKSMWTWAATWSWWSMMATWKR 246

Qy        351 GGCTAATCCTGAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAA 410
              :  : :   ::  :  :::::    :  ::::  |:|: |:  :  |: :|::|:| |:|
Db        247 KMRWGWMWAYSSTWMMRWWWYCCTRRAMYKMKAATMTRCTWYMWMAAMTSAWRAWCAGWA 306

Qy        411 TGAGTTATGTGATACTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGA 470
              |:  |::: :: | : ::  :::|   :  :  :   : |:::::|:|:: | :: ::
Db        307 TSYATKWYAWRTTYM-KYRRRYYACMAWTRWTKMWSMMRTKKWWYTMARSWGCKKWMWAK 365

Qy        471 TGCCATTAAATTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGA 530
               ::    :| |: |: ::  : : :::   | : |::::|::::  :    ::| :
Db        366 ARMATMCMAYTYSTYRYYAWRRRWMRWSWAAWYSAKYWYAKSYYYMYKCARKWWTTYAYY 425

Qy        531 AAAAGCAAAAAAACTATCGAGTGATTT 557
              ||||:  : || :   :: : : ::|:
Db        426 AAAAKATRTAAYMGRTWYTMCWTMYTK 452


RESULT 10
AAA70105
ID   AAA70105 standard; DNA; 5940 BP.
XX
AC   AAA70105;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:238.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US026796.
XX
PR   05-NOV-1998;   98US-0107131P.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection.
XX
PS   Disclosure; Page 460-462; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
```

```
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II). (I) and
CC   (II) are useful for the development of vaccines against P. falciparum
CC   infection. (I) and polyclonal antisera or a monoclonal antibody raised to
CC   immunogens comprising the sequences of (I), are useful in the detection
CC   of infection with P. falciparum. Furthermore, (I) (especially when they
CC   are rifins or secreted or membrane proteins) can aid the identification
CC   of drugs to treat or prevent P. falciparum infection, or they can be used
CC   to identify drug resistance in P. falciparum. Sequencing of the
CC   Plasmodium chromosome 2 and the subsequent identification of proteins
CC   encoded by it will help to expand our understanding of parasite biology,
CC   a process hampered by the complexity of the parasitic lifecycle, and
CC   provide new targets for vaccine and drug development. Parasite resistance
CC   to drugs and mosquito resistance to insecticides have led to a resurgence
CC   of malaria in many parts of the world, and there is a pressing need for
CC   vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
CC   represent nucleotide and protein sequences given in the present
CC   invention, but which are not specifically mentioned within the
CC   specification
XX
SQ   Sequence 5940 BP; 3106 A; 343 C; 879 G; 1612 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 51.8;  DB 3;  Length 5940;
  Best Local Similarity   54.5%;  Pred. No. 0.37;
  Matches  104;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              | |  | ||||||  | ||| |||  ||| |  |||||  |  ||||| |||  |   |
Db        158 AAACTAAATATAAAAATGATAATTCCTTAGATGATAATATTAATGAGGACATAATAAATA 217

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |  ||| | |  ||   ||  |   |  |  ||||     | |
Db        218 ATAATAATAATAATAATAATGATAATAATAATGATAACAATAATGATAATAATAATGATA 277

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | |||  ||||   |||||  |||||   |   || ||||   |
Db        278 ATAATAATGATAATAATAATGATAATAATAATGAGAATAATAATGATAATAATAATTTTA 337

Qy        248 TAAAAAAATGC 258
                ||  |  ||
Db        338 ATAATTATAGC 348


RESULT 11
ABL12834
ID   ABL12834 standard; cDNA; 11143 BP.
XX
AC   ABL12834;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 32984.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB68731.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 32984; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 11143 BP; 3489 A; 2184 C; 2249 G; 3221 T; 0 U; 0 Other;

  Query Match             4.9%;  Score 51.6;  DB 4;  Length 11143;
  Best Local Similarity   52.9%;  Pred. No. 0.46;
  Matches  111;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||  |||||||||  | | |||| | | || ||||||  |  |||    | |  |  | |
Db       4467 ATATAATATATAATATATAATATTATATTAATAATAATATTAATGATAATAATAATAATA 4526

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |  || | |  ||      ||   | ||  ||||    || |
Db       4527 ATAATAATAATAATAATAATAATTATAATAATAACCATAATAATAATAATAATAATAATA 4586

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   |||  |  |||||       || |||| |||
Db       4587 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 4646

Qy        248 TAAAAAAATGCATTAGTTTGACGACAGGTA 277
                || ||    | || |   |  ||   ||
Db       4647 ATAATAATAATAATAATAACAGTACCAATA 4676


RESULT 12
AAA70168
ID   AAA70168 standard; DNA; 7722 BP.
XX
AC   AAA70168;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:301.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US026796.
XX
PR   05-NOV-1998;   98US-0107131P.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
```

PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection.
XX
PS Disclosure; Page 504-506; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
CC (II) are useful for the development of vaccines against P. falciparum
CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
CC immunogens comprising the sequences of (I), are useful in the detection
CC of infection with P. falciparum. Furthermore, (I) (especially when they
CC are rifins or secreted or membrane proteins) can aid the identification
CC of drugs to treat or prevent P. falciparum infection, or they can be used
CC to identify drug resistance in P. falciparum. Sequencing of the
CC Plasmodium chromosome 2 and the subsequent identification of proteins
CC encoded by it will help to expand our understanding of parasite biology,
CC a process hampered by the complexity of the parasitic lifecycle, and
CC provide new targets for vaccine and drug development. Parasite resistance
CC to drugs and mosquito resistance to insecticides have led to a resurgence
CC of malaria in many parts of the world, and there is a pressing need for
CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
CC represent nucleotide and protein sequences given in the present
CC invention, but which are not specifically mentioned within the
CC specification
XX
SQ Sequence 7722 BP; 3627 A; 688 C; 1023 G; 2384 T; 0 U; 0 Other;

Query Match 4.9%; Score 51; DB 3; Length 7722;
Best Local Similarity 51.5%; Pred. No. 0.58;
Matches 117; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

```
Qy         67 GATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGT 126
              |||||  |  || ||    || ||  |   || |||  | || || | || | ||||  |
Db       3931 GATAGTGTTAATGATTCTAATGATGATACCAACAATGTTAATGTTAATGTTAATGTTAAT 3990

Qy        127 GAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAA 186
              ||  || |   |    |    ||||| |  ||   || || |  ||   | |||  ||||
Db       3991 GATTATTATGATGATGATGATGATGATAACAATAATAATAACCGAATAGACAAAAAAAAA 4050

Qy        187 AATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAAT 246
              ||  ||||| ||||    |||  ||||      |   || ||  ||| |  |  | |
Db       4051 AAACATAAAAAAAAGAAATATAATAATGAACCTATGGAAAAAATAGACGTTTCTGATCTA 4110

Qy        247 TTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCAT 293
               || || ||    |||    ||  | ||  ||    || |||   ||
Db       4111 GTAGAAGAAATATTTAACGGGAAAATAGTAAATATATGTATGCATAT 4157
```

RESULT 13
ABD32715
ID ABD32715 standard; DNA; 337344 BP.
XX
AC ABD32715;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated genomic DNA HD14-044.
XX
KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW leukaemia; lymphoma; CAP.
XX
OS Homo sapiens.
XX
PN WO2004074320-A2.
XX
PD 02-SEP-2004.
XX
PF 17-FEB-2004; 2004WO-US004730.
XX
PR 14-FEB-2003; 2003US-00367094.
PR 14-MAR-2003; 2003US-00388838.
PR 15-APR-2003; 2003US-00417375.
PR 13-JUN-2003; 2003US-00461862.
PR 15-SEP-2003; 2003US-00663431.
PR 15-DEC-2003; 2003US-00737318.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Morris DW, Malandro MS;
XX
DR WPI; 2004-652914/63.
XX
PT New isolated cancer-associated polynucleotides and polypeptides useful
PT for diagnosing, preventing or treating cancers, especially lymphoma and
PT leukemia, or in screening for agents that modulate cancer.
XX
PS claim 16; seqid 290; 310pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising at least 10
CC contiguous nucleotides of any of the 233 polynucleotide sequences given
CC in the specification, or its complement. The nucleic acids encode cancer-
CC associated proteins. Also included are an expression vector comprising
CC the isolated nucleic acid cited above, a host cell comprising the above
CC recombinant nucleic acid or expression vector, a microarray for detecting
CC a cancer-associated (CA) nucleic acid comprising at least one probe
CC comprising at least 10 contiguous nucleotides of any of the above-
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC an open reading frame of a CA sequence selected from any of the 95
CC polynucleotide sequences as mentioned in the specification, or its
CC complement), an isolated antibody, (or its antigen binding fragment) that
CC binds to the above polypeptide, a hybridoma that produces the above
CC monoclonal antibody, a pharmaceutical composition comprising the above
CC antibody and a pharmaceutical excipient, a kit for detecting cancer
CC cells(comprising the antibody cited above, methods for diagnosing cancer
CC or for detecting the presence or absence of cancer cells in an
CC individual, a method for inhibiting growth of cancer cells in an
CC individual, a method for delivering a therapeutic agent to cancer cells
CC in an individual, an electronic library comprising the above
CC polynucleotide or polypeptide (or their fragments), methods of screening
CC for anticancer activity or for a bioactive agent capable of modulating
CC the activity of a CA protein (CAP), methods for detecting cancer
CC associated with expression of a polypeptide in a test cell sample, a
CC method for treating cancers and a method for inhibiting the expression of
CC CA gene in a cell. The composition and methods are useful for detecting,
CC diagnosing, preventing and treating cancers, especially lymphoma and
CC leukaemia. These may also be used in screening for agents that modulate
CC cancer. The present sequence is a human CAP genomic sequence. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 337344 BP; 88398 A; 73747 C; 76975 G; 95377 T; 0 U; 2847 Other;

Query Match 4.9%; Score 50.8; DB 13; Length 337344;
Best Local Similarity 46.2%; Pred. No. 1.2;
Matches 169; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

```
Qy          5 ATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTAC 64
              || | | ||| | ||| | ||||   | | |  ||| |  |     | || |   | ||
Db     202599 ATATATATATAAATATATATATAATATATAAATATATATATATAAATATATATATAATAT 202658

Qy         65 TGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTT 124
                 |    ||||| ||  |    || | | |||| |   ||      |      ||  | |
Db     202659 ATAAATATATATACAAATATATATAATATATAAATATATATACAAATATATATATAATAT 202718

Qy        125 GTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAA 184
               | || ||| ||    | | | | || | ||| |      |||  |||  |||| |  |
Db     202719 ATAAATATATATACAAATATATATATAATATATAAATATATATACAAATATATATATAAT 202778

Qy        185 AAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGA 244
              | |  |||| | | |    |      ||   |  || || |||  |     |||    | |
Db     202779 ATATAAATATATATACAAATATATATATAATATATAAATATATATACAAATATATATATA 202838
```

```
Qy        245 ATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATG 304
              || || ||| ||  ||   | |    | |     | | |  ||   ||      |   ||
Db     202839 ATATATAAATATATATACATATATATATAATATATAAATATATATACAAATATATATATA 202898

Qy        305 ATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAAA 364
              || ||| |  | ||  |   ||| | |  |   | |  | |  |  | |||    | |
Db     202899 ATATATAAATATATATACAAATATATATATAATATATAAATATATATACAAATATATATA 202958

Qy        365 TTGTAT 370
              |  |||
Db     202959 TAATAT 202964
```

RESULT 14
AAA70117
ID AAA70117 standard; DNA; 3567 BP.
XX
AC AAA70117;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:250.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US026796.
XX
PR 05-NOV-1998; 98US-0107131P.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection.
XX
PS Disclosure; Page 470-471; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
CC (II) are useful for the development of vaccines against P. falciparum
CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
CC immunogens comprising the sequences of (I), are useful in the detection
CC of infection with P. falciparum. Furthermore, (I) (especially when they
CC are rifins or secreted or membrane proteins) can aid the identification
CC of drugs to treat or prevent P. falciparum infection, or they can be used
CC to identify drug resistance in P. falciparum. Sequencing of the
CC Plasmodium chromosome 2 and the subsequent identification of proteins
CC encoded by it will help to expand our understanding of parasite biology,
CC a process hampered by the complexity of the parasitic lifecycle, and
CC provide new targets for vaccine and drug development. Parasite resistance
CC to drugs and mosquito resistance to insecticides have led to a resurgence
CC of malaria in many parts of the world, and there is a pressing need for
CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
CC represent nucleotide and protein sequences given in the present
CC invention, but which are not specifically mentioned within the
CC specification
XX

```
SQ   Sequence 3567 BP; 1641 A; 269 C; 473 G; 1184 T; 0 U; 0 Other;

  Query Match             4.8%;  Score 50.6;  DB 3;  Length 3567;
  Best Local Similarity   45.5%;  Pred. No. 0.61;
  Matches  218;  Conservative    0;  Mismatches  259;  Indels    2;  Gaps    1;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||  |   |||||    |||||  |    | |||| | ||| | |       |  | |
Db       1382 ATAATAATGATAATATTAATTATAATAATGATAATATTAATTATAATAATGATAATATTA 1441

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||  |    | |  || | |  ||  || ||| |  | ||||  |  || |
Db       1442 ATTATAATAATGATAATATTAATCATAATAATGATAATATTAATTATAATAATGATAATA 1501

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              ||||  |  | |||  | || ||||||  ||| | ||||||  |    ||| ||||  ||
Db       1502 TTAATCATAATAATGATAATATTAATTATAATAATGATAATATTAATTATAATAATGATA 1561

Qy        248 TAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATC 307
              |  || |  | || | |       |          ||   |      || |  ||      |
Db       1562 ATATTAATTATAATAATATTTGTCCGACAGGTGATAAAAATCATATTGAAAAGATCCTTT 1621

Qy        308 TATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAAATTG 367
              ||  | || |      |      | | |  || |||      |||  |||   |  |
Db       1622 TAAAAACACAATCCACAAATGTACAGAATGTTAAAAAGAAAAAGATTAAAGTTTTATACA 1681

Qy        368 TATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGTGATACTG 427
              ||    ||       ||   |  || ||  || |||| ||   |||   | | | |
Db       1682 ATTTAAAAACATTTATTAATTTATTTTTAAACAAATATCCCAAGTGGTTTCATAAAATTA 1741

Qy        428 TTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAATTTTTCT 487
              |||||  |  |      |||  |      | | ||     |||  ||
Db       1742 AAGCAGATTTATTAATCCTGTTCATATTAAAAGGTAATGATTATCTTCCTAAGATTAGAG 1801

Qy        488 TCCGTAGAGTTGG--AGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAAAAAC 544
              | |   |||    || ||| | || |  ||   ||||||  | ||   |||||||
Db       1802 AAGGGAATTTTGGTATCTTTTTTGAAGCATATTTTAAAATGCTTGAGAATATAAAAAAC 1860
```

RESULT 15
ACL35887/c
ID ACL35887 standard; cDNA; 2000 BP.
XX
AC ACL35887;
XX
DT 02-JUN-2005 (first entry)
XX
DE Rice stress-regulated promoter SEQ ID NO:14450.
XX
KW ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW agriculture.
XX
OS Oryza sativa.
XX
PN WO2003008540-A2.
XX
PD 30-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US019668.
XX
PR 22-JUN-2001; 2001US-0300112P.
PR 24-AUG-2001; 2001US-0314662P.
PR 26-SEP-2001; 2001US-0325277P.
PR 21-NOV-2001; 2001US-0332132P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
PI Moughamer T, Provart N, Ricke D, Zhu T;
XX
DR WPI; 2003-248011/24.
XX
PT New stress-responsive nucleic acid, useful for altering the

PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT stress, salt stress or osmotic stress.
XX
PS Claim 48; SEQ ID NO 14450; 89pp; English.
XX
CC The invention relates to novel abiotic stress responsive polynucleotides
CC and polypeptides. Also disclosed are vectors, expression cassettes, host
CC cells, and plants containing such polynucleotides. Also disclosed are
CC methods for using the polynucleotides and polypeptides to alter the
CC responsiveness of a plant to abiotic stress. The invention is useful in
CC agriculture. The nucleic acid is useful for determining whether a test
CC plant has been exposed to an abiotic stress condition. It is also useful
CC for selecting an agent that alters abiotic stress regulated
CC polynucleotide expression in a plant cell, and to identify a homolog or
CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC molecule and the polypeptide encoded by it are useful in altering the
CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC stress, osmotic stress or any of their combinations. The present sequence
CC is used in the exemplification of the invention
XX
SQ Sequence 2000 BP; 410 A; 329 C; 253 G; 416 T; 0 U; 592 Other;

Query Match 4.8%; Score 50.4; DB 11; Length 2000;
Best Local Similarity 15.3%; Pred. No. 0.61;
Matches 131; Conservative 314; Mismatches 400; Indels 9; Gaps 2;

```
Qy        154 ATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAAT 213
              ||: :   ::: ::: :| ::||:|:: :::   ::|::: :    :| :|::  :|: :
Db        936 ATMTMAMWWMYSRWMYRATRMTAWAMWTMWWCTCRWARYMKWASSKRTGYTWRRWWTWYW 877

Qy        214 TTCAATGAAGATAAT--TTAGGCTATGATAAGAATTTAAAAAAATGCATTAGTTTGACGA 271
               :::  ::| :   |  ::  : : :: :  ||  |::|:| :  :|: :    ||:
Db        876 CYSRTMKWAWWAKGTGYKWTTRRYTWKYKSYGATSTWWAWATRTARCWAWKWRCTGMGTG 817

Qy        272 CAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTAGCAGATGGAGCGTTATCAA 331
              : |  |:  |  :: :  : : ||::::   :: :::::    : :        :::::
Db        816 KTKATKTAWTATAWRKAWMAMARTGCWRMKKGWWYAWWWWWTWYWRGKATWAYMMWKYMR 757

Qy        332 AAATAGTGAAAGTTTTGAAGGCTAATCCTGAAATTGTATTGGCTACGCGAGCGTATGGTT 391
              :   : :::|::|||       :|  : :  : :   :: : : :: :: :  :   : :
Db        756 MTSRMMYRWARKTTTCYKGWATKATMMAKWSRSKSMSMYGKMSGWYYSSTKWMWGATRMW 697

Qy        392 GGTTTAAGGAAAATCCGAATGAGTTATGTGATACTGTTCGTCATTTAACAGACGATACTT 451
              | ::    ::   ::   :::  | |  :|:| ::  :        |:  : :::|::|
Db        696 GMKWGKGCRRCGGYMKWSMKSKWTCASAWGMTKYKYSYRCRGCGGTWTTRWWMRAWMCMA 637

Qy        452 TATTTCAGCCGGGGGCTGATGCCATTAAATTTTTCTTCCGTAGAGTTGGAGTTATTTCAG 511
              ::|:|   :  :| : :: | : : :::  :::| |   :::    | |  :    :  :
Db        636 KWTKTGSYYWCRGYSKYKTTWYWRAWWWSCWKKTGTGRTKYMWGTSTYGTTKMYSAYWKR 577

Qy        512 GCTTTATTGTCAATGCTGAAAAAGCAAAAAAACTATCGAGTGATTTATTTGATGGGCGTT 571
              :::: :|  ||| : :     : | : ||: :::: ||:  | ::  | :: :|:|:
Db        576 MWSKKRWGRTKYATGWKYWGYTGWYCAYWSAAMCWKMRKGTKTRTGRKSRGCWRTKCKTY 517

Qy        572 TATATTATCAAATGTACCTTGCTGGTATGCTAATGGCTGA-------AGGTCAGGGATAC 624
              ::   :: ::::|||     |:   ::   :::::    |       ::| :: :  | :
Db        516 WRMYRWMGYRWRTGTSRGRGGMAATKMSTKYRWKRWWRMAWATKSWKWRGMMWTKWCTKY 457

Qy        625 TATTTTAGCGACGTGATGACATTGTCGAGGGATACAGAGGCTCCTGACTTTGGTAACGCT 684
              : :::: :|:|:: |  | |:|:|       |: ||: ||   :  : | : : :::|
Db        456 AYKRWYMRAKAYGKKTTWGCSTYGKCTTWAYTRASYGAKWCTRTGKTTKGTCRGMKYRMT 397

Qy        685 GGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTATAAACCAGAGGGCCGTATACATATG 744
              |  :    |::  ||:|:: : :|:: : :|       :: :|  : :  ::
Db        396 TAAGTKAKYYKAKRKTTTKTMWKMRYGSKWRWKTTAGYTRAKWKTYKYTRTTSWWRSKWK 337

Qy        745 GTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAAGATACAACAAAAATTGATGGCGTT 804
              :|: : :   |: |     ::: :|::: :  ::   ::      :   ::   : ||:
Db        336 KTWWRKKWWRTKGTYKCTGKWRTMAMWWKKTARRTSWWKTKKYTTMKGKRWKCAMRAGTK 277

Qy        805 TATGCTGGAATTAGAAAAGACTTAGCGAACTATTTTTATCCTTATATTCGAGATCAACTC 864
              : ||: : :   | : :     : :  :::: |  |:|  :|:|:  |:  |
Db        276 AKWTCTRWWYYMYWAKRTRTGMMCKRSKGMWWWCTGAAYCACWAYAKATGWYCTKYGRCA 217

Qy        865 GACTTGCCTCTTTATACTTATATTAAAATGATAAATAAATTTCGGAAAATGGGATTTTCA 924
              ||:|   : |: | : ::   :| | : :  :  : : ::   ::|:| : |||: ::|
Db        216 MACWTTGTYKTWSARWWYWTVKWTGAYWSKCAWTSWYWYWKMWWRMAWAGRTGATWRWMA 157

Qy        925 AATGAAAAGCTTTTCTATGTGCATGCCTTTTTAGGGTATGTACTAAAACGGAGGGGCTAT 984
              | :: | ::|:|::::: |||          ||  : : :  : :| :  : ::     ||:
Db        156 ATYRCACRKCWTKWMWMGGTGAYAAWAAATTRKRMKRMATWTMTYWKKMYRMWWWATTAY 97

Qy        985 GATGCTTTAATTAA 998
              ::: : :  ||:::
Db         96 KWYAMAYAYATWWW 83
```

Search completed: May 11, 2006, 09:42:43
Job time : 477.701 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 09:43:19 ; Search time 3252.76 Seconds
(without alignments)
15016.694 Million cell updates/sec

Title: US-09-423-093B-3_COPY_3977_5020
Perfect score: 1044
Sequence: 1 atgaatgtttttatcagtat.....ctccgcgtcttggtatttaa 1044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	67.4	6.5	1204	10	CNS016E2	AL106628 Drosophil
c	2	64	6.1	872	9	BZ385991	BZ385991 EINAJ93TR
	3	61	5.8	579	1	AU038989	AU038989 AU038989
c	4	60.4	5.8	1101	10	CNS0039G	AL063921 Drosophil
	5	59.4	5.7	520	9	AQ656286	AQ656286 Sheared D
c	6	58.8	5.6	519	3	BJ440670	BJ440670 BJ440670
c	7	58.6	5.6	952	10	CNS01016	AL098388 Drosophil
	8	58.4	5.6	1101	10	CNS00EO7	AL069440 Drosophil
	9	58	5.6	844	10	BX139987	BX139987 Danio rer
c	10	57.8	5.5	705	3	BJ446775	BJ446775 BJ446775
	11	57.6	5.5	1101	10	CNS0106X	AL098595 Drosophil
	12	57.2	5.5	1101	10	CNS00LT2	AL078714 Drosophil
c	13	57	5.5	305	3	BJ442567	BJ442567 BJ442567
	14	56.6	5.4	429	10	CG803197	CG803197 1118039H0
c	15	56.6	5.4	430	10	CG803196	CG803196 1118039H0
	16	56.2	5.4	595	3	BM273696	BM273696 PfESToaa5
	17	56.2	5.4	1043	10	CNS0145P	AL103735 Drosophil
c	18	56.2	5.4	1101	10	CNS00G0M	AL071284 Drosophil
	19	56	5.4	299	1	AU038406	AU038406 AU038406
c	20	56	5.4	661	10	CNS02QVJ	AL209800 Tetraodon
c	21	55.8	5.3	822	7	CK472148	CK472148 AGENCOURT
c	22	55.6	5.3	401	9	AQ639568	AQ639568 927P1-4D3
	23	55.4	5.3	469	1	AU076358	AU076358 AU076358
	24	55.2	5.3	311	1	AU054171	AU054171 AU054171
	25	55.2	5.3	410	1	AU037570	AU037570 AU037570
c	26	55.2	5.3	610	8	DN469310	DN469310 USDA-FP_1
	27	55.2	5.3	822	9	BZ430021	BZ430021 BONEU51TR
c	28	55.2	5.3	1201	10	CNS007E3	AL067050 Drosophil
	29	55	5.3	471	10	CG807194	CG807194 1118078H0
c	30	55	5.3	670	9	BZ388480	BZ388480 EINDL21TF
c	31	55	5.3	829	10	BX173672	BX173672 Danio rer
	32	55	5.3	1092	10	CNS020K7	AL175696 Tetraodon
	33	54.8	5.2	650	9	BZ423232	BZ423232 id46h10.g
c	34	54.8	5.2	700	9	BZ421879	BZ421879 id46h10.b
	35	54.8	5.2	832	9	BH391984	BH391984 AG-ND-138
c	36	54.8	5.2	891	10	CW952279	CW952279 TcB38.1_B
c	37	54.8	5.2	984	10	CZ934526	CZ934526 253530 To
	38	54.6	5.2	593	9	BH657010	BH657010 BOHUQ52TR
	39	54.6	5.2	613	3	BJ425106	BJ425106 BJ425106
	40	54.6	5.2	905	9	BH138573	BH138573 ENTPB03TR
c	41	54.4	5.2	600	11	DR26K22T	AL976955 Danio rer
c	42	54.2	5.2	862	10	DU065131	DU065131 72513 Tom
	43	54.2	5.2	1101	10	CNS017KE	AL108152 Drosophil
c	44	54	5.2	736	9	CC077693	CC077693 CSU-K33r.
c	45	54	5.2	928	10	CNS00DKY	AL071865 Drosophil

ALIGNMENTS

RESULT 1
CNS016E2/c
LOCUS CNS016E2 1204 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106628
VERSION AL106628.1 GI:5622852
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1204)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1204
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN15A12"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"
ORIGIN

Query Match 6.5%; Score 67.4; DB 10; Length 1204;
Best Local Similarity 24.5%; Pred. No. 0.00075;
Matches 110; Conservative 159; Mismatches 176; Indels 4; Gaps 1;

Qy 95 TAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCCACAGAGAGATGAGA 154

```
Db       1192 WATAKWAKWAWKATAAKAWWKAKAKWKAAAAAAAAAAAAAAAAAAKAAAADAAAAKKKAAA 1133

Qy        155 TAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAATT 214

Db       1132 AAAAAKAADDAAADAAAAAKAKAAKAKAAAAAAADAAWKKKAWAAAAAAKWAAAKWKATK 1073

Qy        215 TCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCATTAGTTTGACGACAG 274

Db       1072 KDDDDKDKAAADAKKKKKKAKADADKAADKDKDKKAKADAWDDADDAAKKKKKAKKKDKK 1013

Qy        275 GTAAATATTGCATGATCATGGGCAACGATGATCTATTAGCAGATGGAGCGTTATCAAAAA 334

Db       1012 KDDA----TDADWAAKAAKKKDKKAWKKKKDKKKAKKAKAWKKKDKKDKKDKKKKKKDK 957

Qy        335 TAGTGAAAGTTTTGAAGGCTAATCCTGAAATTGTATTGGCTACGCGAGCGTATGGTTGGT 394

Db        956 KAKKKKKAAAKDAAAAAAAAAKKKDKKKDAAKAAAKKKKKDAAKKKAKKKDDAWAADAD 897

Qy        395 TTAAGGAAAATCCGAATGAGTTATGTGATACTGTTCGTCATTTAACAGACGATACTTTAT 454

Db        896 KAAAKKKAAAKKKDKKKDAADKKAAKDAAKKKKDKAAKAADKDAAKKWDKKKKAGGAARA 837

Qy        455 TTCAGCCGGGGGCTGATGCCATTAAATTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCT 514

Db        836 AAKKGKKGGGRKGGGAGGGRAAAAGATWTTWAKTDTTKTKAGDAKTTTTKKTTTWATKAG 777

Qy        515 TTATTGTCAATGCTGAAAAAGCAAAAAAA 543

Db        776 AKWAKKTRAAAKDARGKDAATTAKAARAA 748
```

RESULT 2
BZ385991/c
LOCUS BZ385991 872 bp DNA linear GSS 30-APR-2003
DEFINITION EINAJ93TR EI_10_12_KB Entamoeba invadens genomic clone EINAJ93, genomic survey sequence.
ACCESSION BZ385991
VERSION BZ385991.1 GI:30231873
KEYWORDS GSS.
SOURCE Entamoeba invadens
ORGANISM Entamoeba invadens
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 872)
AUTHORS Wang,Z., Samuelson,J., Clark,C.G., Eichinger,D., Paul,J., van Dellen,K., Hall,N., Anderson,I. and Loftus,B.
TITLE Gene discovery in the Entamoeba invadens genome
JOURNAL Mol. Biochem. Parasitol. 129 (1), 23-31 (2003)
PUBMED 12798503
COMMENT Other_GSSs: EINAJ93TF
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
DNA was provided by Daniel Eichinger
Seq primer: TR
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .872
/organism="Entamoeba invadens"
/mol_type="genomic DNA"
/strain="IP-1"
/db_xref="taxon:33085"
/clone="EINAJ93"
/clone_lib="EI_10_12_KB"
/note="Vector: pHOS2; Site_1: BstXI; Total genomic DNA was isolated from early log phase trophozoites of E. invadens IP-1 using a Qiagen plant DNA extraction kit. A shotgun medium-size plasmid library (average insert size of 10 - 12 kb) was generated by random mechanical shearing of E. invadens genomic DNA, repairing the ends of DNA fragments with T4 Polymerase, adding BstXI adaptors and ligating into the BstXI site of a pUC-derived vector pHOS2."
ORIGIN

Query Match 6.1%; Score 64; DB 9; Length 872;
Best Local Similarity 53.6%; Pred. No. 0.0039;
Matches 133; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

```
Qy         81 TCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCC 140

Db        835 TCCAAATTTATAGTAATATATTTGAAATAAAAATGTTTTATTTTGTTAAGGGAAAAATAT 776

Qy        141 ACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAA 200

Db        775 ATATATTTATATAAAAAATTAAATTAAAAATAAGTAATTTTTAAAAATTTATCAAAAAAA 716

Qy        201 TCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCAT 260

Db        715 TTGAAAAGTAAATATAATAAAAGATATAATAAAATATGATGTGAAAATAAAAAACTGAAT 656

Qy        261 TAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTAGCAGATGG 320

Db        655 TATTAAATATAAATACACCAATTGGAATGAAATGCACAGTGTTGATTTGTTATGAAATGA 596

Qy        321 AGCGTTAT 328

Db        595 ACACTTGT 588
```

RESULT 3
AU038989
LOCUS AU038989 579 bp mRNA linear EST 29-MAR-1999
DEFINITION AU038989 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium discoideum cDNA clone SSM246, mRNA sequence.
ACCESSION AU038989
VERSION AU038989.1 GI:3985742
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 579)
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
TITLE The Dictyostelium developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development
JOURNAL DNA Res. 5 (6), 335-340 (1998)
PUBMED 10048482
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES Location/Qualifiers
source 1. .579
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSM246"
/dev_stage="slug"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN

Query Match 5.8%; Score 61; DB 1; Length 579;
Best Local Similarity 49.8%; Pred. No. 0.017;
Matches 154; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

```
Qy          4 AATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTA 63
```

```
Db         32 AATAATTATCATATAAATTCAAATACTTCAATTATTAATGGTCATATTTTTGATGAAAAT 91

Qy         64 CTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTT 123

Db         92 ATTAATAATTATCATAGTAATAATAATAATAATGACGATGATGATGATGATAAAAATGAA 151

Qy        124 TGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCA 183

Db        152 GAAGATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 211

Qy        184 AAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAG 243

Db        212 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 271

Qy        244 AATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGAT 303

Db        272 AATAATAATAATGAAATTATTAAAAATATAATTTTAAATTGGATTAAACAATGTAATGAT 331

Qy        304 GATCTATTA 312

Db        332 AATCAATCA 340


RESULT 4
CNS0039G/c
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.8%;  Score 60.4;  DB 10;  Length 1101;
  Best Local Similarity   17.6%;  Pred. No. 0.023;
  Matches   90;  Conservative  214;  Mismatches  206;  Indels    0;  Gaps    0;

Qy         77 ATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAAT 136

Db        950 AKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDKDDDGKGKDADDDTDGTKDDDD 891

Qy        137 CTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAAC 196

Db        890 KDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTWDAAADDWWADDRWDAWAWKWD 831

Qy        197 AAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAAT 256

Db        830 DAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRDAADDRDDAATWTTWTTTTTRD 771

Qy        257 GCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTAGCAG 316

Db        770 TDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTWKRRWKRRDTRWDDADADD 711

Qy        317 ATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAAATTGTATTGGCTA 376

Db        710 TARDDRRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDAWWADAAWWTTTDTDTDDWDKR 651

Qy        377 CGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGTGATACTGTTCGTCATT 436

Db        650 DRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRADRWDRWAADTWTDARKADRDWA 591

Qy        437 TAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAATTTTTCTTCCGTAGAG 496

Db        590 KARAWRARRDRARAARADRRWTTKGKTTTATWTTWAARAAWWAWWAWATTTATWTTTWTT 531

Qy        497 TTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAAAAACTATCGAGTGATT 556

Db        530 TWTTTTTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWAAAAAAAWAATTTTTTTTTTT 471

Qy        557 TATTTGATGGGCGTTTATATTATCAAATGT 586

Db        470 WAWTAAAWWTAWTWTTTTWTTTTWWAATTT 441


RESULT 5
AQ656286
LOCUS       AQ656286                520 bp    DNA     linear   GSS 23-JUN-1999
DEFINITION  Sheared DNA-27K23.TR Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-27K23, genomic survey sequence.
ACCESSION   AQ656286
VERSION     AQ656286.1  GI:5164054
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
            Donelson,J., Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: Sheared DNA-27K23.TF
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Reverse
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .520
                     /organism="Trypanosoma brucei"
```

```
                     /mol_type="genomic DNA"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-27K23"
                     /clone_lib="Sheared DNA"
                     /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from a cloned population of
                     Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                     sheared to give a tight size distribution (approx 2 kb).
                     The v + i method used for the library construction is
                     described in detail in Smith, H.O. and Venter, J.C.
                     (Making small insert libraries for whole genome shotgun
                     sequencing projects.  In Genome Sequencing: A Practical
                     Approach, eds. M. Vaudin and B. Barell, Oxford University
                     Press, 1999)."
ORIGIN

  Query Match             5.7%;  Score 59.4;  DB 9;  Length 520;
  Best Local Similarity   51.3%;  Pred. No. 0.036;
  Matches  138;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||  |   |||||  |  || ||  |   || ||||||  ||  | |   |  |  | |
Db         30 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 89

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |   || | |  ||   || ||   | ||  ||||    || |
Db         90 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 149

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| |  |||||       || |||| |||
Db        150 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 209

Qy        248 TAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATC 307
              || ||    | || |   |  |    |||  ||   |  |  |     || ||| ||
Db        210 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGATAATA 269

Qy        308 TATTAGCAGATGGAGCGTTATCAAAAATA 336
              |  ||  ||| ||| || | || | |||
Db        270 GAAAAGTGGATTGAGGGTGAGCATATATA 298


RESULT 6
BJ440670/c
LOCUS       BJ440670                519 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ440670 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv44k10 3', mRNA sequence.
ACCESSION   BJ440670
VERSION     BJ440670.1  GI:19415392
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .519
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv44k10"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

  Query Match             5.6%;  Score 58.8;  DB 3;  Length 519;
  Best Local Similarity   55.5%;  Pred. No. 0.048;
  Matches  111;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy         77 ATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAAT 136
              ||||| |  || ||  |   || |||| | || ||      | |  | ||||   |||
Db        264 ATAATAATAATAATGATGATAATAATAGTAATGTTNTNAATAATAATAGTGATTTTAATG 205

Qy        137 CTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAAC 196
              |    | |   || | |  ||   || ||   | ||  ||||    || ||||||||
Db        204 ATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 145

Qy        197 AAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAAT 256
              | |||  | ||  ||||   |||||| ||||| |||   || |||| ||||| || ||
Db        144 ATAATAATAATAATAATAATAATGAAAATAATATAGAAGATAATAATAATTTTAATAATA 85

Qy        257 GCATTAGTTTGACGACAGGT 276
              | || |   ||   || |
Db         84 ATAATAATAATACTGAAGTT 65


RESULT 7
CNS01016/c
LOCUS       CNS01016                952 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03B09 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098388
VERSION     AL098388.1  GI:5609999
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 952)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .952
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN03B09"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             5.6%;  Score 58.6;  DB 10;  Length 952;
  Best Local Similarity   31.3%;  Pred. No. 0.055;
  Matches  101;  Conservative   81;  Mismatches  141;  Indels    0;  Gaps    0;

Qy         22 TGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATATATAAT 81
              |:||     | : |: :||::  |:  | | :| |:  : |:   ::::: |:|:|  ||
Db        840 TRTARKAGTTAKAAKTRTAKRTRTRTATATAKATADTRWATRTRAKWKWRTAKAKAGGAT 781
```

```
Qy          82 CAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCCA 141
Db         780 ATRTAKDKRTGTGTRTTGARTRTTGTAKWTGWGWGTATRTADDDGKKATAKARTRWAAAA 721

Qy         142 CAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAAT 201
Db         720 AARARKATAKATATRTARTATAGTRRKRRKRTAKWDTRTAKATAKATADTATAKRDAKAK 661

Qy         202 CTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCATT 261
Db         660 DKATATTATRTATWWATRTRTGKKATGTRKADATATRKTRTGTATAKAKTRTRAKKATTG 601

Qy         262 AGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTAGCAGATGGA 321
Db         600 TRTGTGWGTGWGRGTAGAKATTRRTATATAAATAGWATRDATAKTRTRTATDTATAWWKT 541

Qy         322 GCGTTATCAAAAATAGTGAAAGT 344
Db         540 KAKDTRDDTATADTRATTAWADD 518


RESULT 8
CNS00EO7
LOCUS       CNS00EO7                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069440
VERSION     AL069440.1  GI:4949583
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29P01"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.6%;  Score 58.4;  DB 10;  Length 1101;
  Best Local Similarity   35.8%;  Pred. No. 0.061;
  Matches  170;  Conservative   68;  Mismatches  236;  Indels    1;  Gaps    1;

Qy          73 ATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGAT 132
Db         546 ATAAATAAWWWATKTTTTTATWWWATATAAAAAWAATTTAATTTTTWATTTTATATATTW 605

Qy         133 AAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAAT 192
Db         606 AAWGWWAAAAAAAAWWTATAAWHTWTAAATTTATAWTTWAWATTTTAAATAWATTTATTTA 665

Qy         193 AAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAA 252
Db         666 AWTTAWWTWTTTAAWAWTTTAAWTATTAAARWWAATTWAAAAWAAAAWTTTWAATAAWAA 725

Qy         253 AAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTA 312
Db         726 ATAAAAATAAATTT-TTTTTATTTAAWWAWTTAAAATTAWTWATWAAWAWTTWTAWAAAA 784

Qy         313 GCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAAATTGTATTG 372
Db         785 AAAATNWTATAAWATTWWKAATAATWTATAWWTTTAWATATWATRTAAATTWTAATWYTT 844

Qy         373 GCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGTGATACTGTTCGT 432
Db         845 TTTWTTTAWTTTAAWATTTTTTTAWWTAAWAWTGTAATWAAATTAWWTTAAAAAWAATGW 904

Qy         433 CATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAATTTTTCTTCCGT 492
Db         905 TTTWTAAATTWWGTTAWTTTTTATTTTWAAATTGTATATAAAAAAWWATTATGTATAWWA 964

Qy         493 AGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAAAAACTAT 547
Db         965 AATRTWWTGTATTKAAAATATWGRATATTKAAAAAAAAAATAWTAAAAWARTKKT 1019


RESULT 9
BX139987
LOCUS       BX139987                844 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  Danio rerio genomic clone DKEY-99E7, genomic survey sequence.
ACCESSION   BX139987
VERSION     BX139987.1  GI:27971314
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 99E7. 99E7 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-99E7"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.6%;  Score 58;  DB 10;  Length 844;
  Best Local Similarity   49.3%;  Pred. No. 0.073;
  Matches  151;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy           1 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 60
Db         344 ATTAATAATAATAATAATGATAATAATAATGATAATAATAATAATGATGATAATAATAAT 403

Qy          61 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 120
Db         404 AATATTGATAATAATGATAATAATAATAATAATGATAATAATAATAATATTGATAATAAT 463
```

```
Qy        121 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 180
Db        464 GATAATAATAATAATAATGATAATAATAATAATATTGATAATATTGATGATAATAATAAT 523

Qy        181 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 240
Db        524 AATGATAATAATAATAATAATGATAATAATAATGATAATAATAATAATAATGATAATAAT 583

Qy        241 AAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAAC 300
Db        584 AATAATGATGATAATAATAATGATGATAATAATGATAATAATAACAATAATAATAATAAT 643

Qy        301 GATGAT 306
Db        644 GATAAT 649
```

RESULT 10
BJ446775/c
LOCUS BJ446775 705 bp mRNA linear EST 13-MAR-2002
DEFINITION BJ446775 Dictyostelium discoideum cDNA library, VF Dictyostelium discoideum cDNA clone ddv63f20 3', mRNA sequence.
ACCESSION BJ446775
VERSION BJ446775.1 GI:19421494
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 705)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .705
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv63f20"
/sex="mat A"
/dev_stage="Growth phase"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Query Match 5.5%; Score 57.8; DB 3; Length 705;
Best Local Similarity 55.0%; Pred. No. 0.08;
Matches 110; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```
Qy         77 ATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAAT 136
Db        277 ATAATAATAATAATGATGATAATAATAGTAATGTTNTNAATAATAATAGTGNTTTTAATG 218

Qy        137 CTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAAC 196
Db        217 ATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 158

Qy        197 AAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAAT 256
Db        157 ATAATAATAATAATAATAATAATGAAAATAATATAGAAGATAATAATAATTTTAATAATA 98

Qy        257 GCATTAGTTTGACGACAGGT 276
Db         97 ATAATAATAATACTGAAGTT 78
```

RESULT 11
CNS0106X
LOCUS CNS0106X 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL098595
VERSION AL098595.1 GI:5610206
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN03K20"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"
ORIGIN

Query Match 5.5%; Score 57.6; DB 10; Length 1101;
Best Local Similarity 19.5%; Pred. No. 0.09;
Matches 65; Conservative 147; Mismatches 122; Indels 0; Gaps 0;

```
Qy         19 ATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATATAT 78
Db        633 ATKTAKRTADTKDTRTAATATAKATKKKARKDKDTAKDAWKWDKATKAKAKKKAAKAKAK 692

Qy         79 AATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCT 138
Db        693 AAAKTATGKDARKAKAKAKAKATKRAKARAWDTATWTATAADAAADKGAKDKAKAKAKAD 752

Qy        139 CCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAA 198
Db        753 ARDKRRWDAKKDKRKRAKAKAAKAADDADADAKDAKAADAKAKADAADADDGRGDGKKKR 812

Qy        199 AATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGC 258
Db        813 AKDRKKKKKKDKAWGDKKKAKKDKAAAKAKADAAGDAKAKRRRAGDKDKAKADAKAAKKA 872

Qy        259 ATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTAGCAGAT 318
Db        873 KAKDDDDAKAATKAKKATKAKKDKAKAKKKKKKKDKDAKAKAKAKDAKDKDDKDDDKDDK 932

Qy        319 GGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGG 352
Db        933 KDAKADKKKWKKDRAKDKAKKDDDKKDKAKDKDK 966
```

RESULT 12
CNS00LT2
LOCUS CNS00LT2 1101 bp DNA linear GSS 14-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:

```
            BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL078714
VERSION     AL078714.1  GI:5102004
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR48P19"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.5%;  Score 57.2;  DB 10;  Length 1101;
  Best Local Similarity   19.5%;  Pred. No. 0.11;
  Matches  117;  Conservative  203;  Mismatches  280;  Indels    0;  Gaps    0;

Qy         65 TGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTT 124
              :::||   :| | :||:  : |:::|| || :     :| | || ||    |  :| ::
Db        498 WKKATTTTWTTTTDAADDKWAAWWDTTTTTAWTTWTTWATTAATATTTTTTTTKWTTWKA 557

Qy        125 GTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAA 184
               : ::  :|    :  | | :    |   ||:   | |:|||:  :|:  :|:: |   :
Db        558 TKTDWTKKAKTAAKTAAWAAWWTTTTTTAATWTTTTKTWTTAKKTWAWKWWAWDTATTTW 617

Qy        185 AAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGA 244
               : |  |:     :|  ||| : ||:  :    |     || |:|    | |      |
Db        618 TTWTTTTTWTTTTTWAATTTTTWTTTWTAWATTTTTTTTTTTAKTKTTTTTTTTTTATTAA 677

Qy        245 ATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATG 304
              ||  ||||:  |: :  | :||   :      |: : : : ::||: ||: :  : :|
Db        678 ATAAAAAAWTTTDTWAAAWWTTTKKKKKKAAAADKWKDAKKWDGAKKATKTKKKDKKAAW 737

Qy        305 ATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAAA 364
              |  : ::::  | ::| | : :         : : :  :|:: :: ||: :| :  ::
Db        738 AAADKKDRKKGGKKKGKGKKGKKGGKKKKKKKKGKGGWGKKGKKKAGDDDAKDKTKKKKK 797

Qy        365 TTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGTGATA 424
              :  | ||:: :  : : :  : ::  :: ::| :: ||||   :  :: |:: ::|: :
Db        798 KAATTTTKKKKGKKKKKKAAKKKAADRTKKTKWDAAAAAAAAAKKTKDKGKKKKKTKKKT 857

Qy        425 CTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAATTTT 484
               :::: ::  :     : | : :   |: ::   |  :|||: :|  |  :::||| :::
Db        858 TKKKKKKKKKKGGGKKDAAAKKKKKGTKKKGKKGKGKGGGKGKGKGGKKDDDAAAAKKK 917

Qy        485 TCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAAAAAC 544
              ::  |: | |:::: |:: : :  :: |:: ::::   :: : |    : :|||:||
Db        918 GKKKKKGKGGGGKKKKGGKKGKGKGGKKKTKKTKKKKKKKKKADAAKGKTKKRAAADAAA 977

Qy        545 TATCGAGTGATTTATTTGATGGGCGTTTATATTATCAAATGTACCTTGCTGGTATGCTAA 604
               :: :| |: ::: :::: : :  ::: | | | :    :::   ::  :::: :: :|:
Db        978 ADWTDAATKKKKKKKKDKKKTKTKKKDAAAAATKKKKKKKKKTKKKKTKKKKKKKKKDAD 1037

Qy        605 TGGCTGAAGGTCAGGGATACTATTTTAGCGACGTGATGACATTGTCGAGGGATACAGAGG 664
              ::: |   ::   :||    :  ::: : :  ::: ::   :: | : |:: :  | :::
Db       1038 KKKKTTGKKDGKGKGGKGKGKKGKKKTKTKKKKKKKKKKTKKKTTADKGKKKKKKAADDK 1097


RESULT 13
BJ442567/c
LOCUS       BJ442567                305 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ442567 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv50d12 3', mRNA sequence.
ACCESSION   BJ442567
VERSION     BJ442567.1  GI:19417289
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 305)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .305
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv50d12"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

  Query Match             5.5%;  Score 57;  DB 3;  Length 305;
  Best Local Similarity   54.0%;  Pred. No. 0.11;
  Matches  108;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy         77 ATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAAT 136
              ||||| |  || ||  |    | |||| | ||  |       | |  | || |   |||
Db        267 ATAATAATAATAATGATGATNATAATAGTAATGGTNTTNATAATAATAGTNATNNTAATG 208

Qy        137 CTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAAC 196
               |    | |   || | |  ||   || ||   | ||  ||||    || ||||||||
Db        207 ATGATNATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 148

Qy        197 AAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAAT 256
              | |||  | ||  ||||   |||||| ||||| |||   || |||| ||||| || ||
Db        147 ATAATAATAATAATAATAATAATGAAAATAATATAGAAGATAATAATAATTTTAATAATA 88

Qy        257 GCATTAGTTTGACGACAGGT 276
              | || |   ||   || |
Db         87 ATAATAATAATACTGAAGTT 68


RESULT 14
CG803197
LOCUS       CG803197                429 bp    DNA     linear   GSS 10-NOV-2003
DEFINITION  1118039H04.y1 1118 - RescueMu Grid S Zea mays genomic, genomic
```

```
            survey sequence.
ACCESSION   CG803197
VERSION     CG803197.1  GI:38238983
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 1118039  row: 2
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .429
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="mixed background W23/A188/B73"
                     /db_xref="taxon:4577"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="1118 - RescueMu Grid S"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units. For more information on RescueMu, go to the web
                     site 'www.zmdb.iastate.edu' and follow the links for
                     'RescueMu.' Grid S was grown at San Diego in 2002. DNA was
                     extracted from leaf strips, double digested using BamHI
                     and BglII, and ligated to form circular plasmids. DH10B
                     cells were transformed and then screened on LB plates with
                     ampicillin."
ORIGIN

  Query Match             5.4%;  Score 56.6;  DB 10;  Length 429;
  Best Local Similarity   52.3%;  Pred. No. 0.14;
  Matches  125;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||   |  |||||| |  ||||| |||   ||  ||||| || || |     | |  || |
Db        134 ATAATAATTATAATTATAATTATAGTTATAATTATAATTATATTTATAATAATAATTATA 193

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              | |||| | |     | |   || | |  || | || ||    |  |  || |     || |
Db        194 ATTATAATTATAATTATAATAATAATAATAATTATAATAATAATTATAATTATAATAATA 253

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| |  |||||    |   || |||| |||
Db        254 ATAATAATAATAATTATAATAATAATAATAATAATAATAATGATAATAATAATAATA 313

Qy        248 TAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGAT 306
              || ||    | ||| |   | |    |||  ||   |  |  |     || || ||
Db        314 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 372


RESULT 15
CG803196/c
LOCUS       CG803196               430 bp    DNA     linear   GSS 10-NOV-2003
DEFINITION  1118039H04.1EL_x1 1118 - RescueMu Grid S Zea mays genomic, genomic
            survey sequence.
ACCESSION   CG803196
VERSION     CG803196.1  GI:38238982
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Very probable ligation site of ends cut by single endonuclease.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 1118039  row: 2
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="mixed background W23/A188/B73"
                     /db_xref="taxon:4577"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="1118 - RescueMu Grid S"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units. For more information on RescueMu, go to the web
                     site 'www.zmdb.iastate.edu' and follow the links for
                     'RescueMu.' Grid S was grown at San Diego in 2002. DNA was
                     extracted from leaf strips, double digested using BamHI
                     and BglII, and ligated to form circular plasmids. DH10B
                     cells were transformed and then screened on LB plates with
                     ampicillin."
ORIGIN

  Query Match             5.4%;  Score 56.6;  DB 10;  Length 430;
  Best Local Similarity   50.2%;  Pred. No. 0.14;
  Matches  140;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy          4 AATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTA 63
              ||| |||||||   || || ||       ||||||||     | |    ||   |
Db        294 AATATTTTTATTTATAATTATAGGGATAGTTATAATAATAATTATAATTATAATTATAAT 235

Qy         64 CTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTT 123
                 ||    | |||||| |  || ||  ||  || |||||| ||  | |        | |
Db        234 AATAATTATAATTATAATAATAATAATAATTATAATAATAATAATAATAATAATTATAAT 175

Qy        124 TGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCA 183
                | |  ||||    |     | |   || |    ||   || ||   | ||  ||||
Db        174 AATAATAATAATAATAATAATAATAATAATTATAATAATAATAATAATAATAATAATAAT 115

Qy        184 AAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAG 243
              || |||||||  | |||  | ||  |||||  ||| |  |||||      ||||||  |
Db        114 AATAATAATAATAATAATAATAATAATTATAATAATAATAATAATTATGATTAT 55

Qy        244 AATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATAT 282
              |||   ||| |   | || |   |  |    |||  ||
Db         54 AATAATAAATATAATAATAATAATAATAATAATAATAAT 16


Search completed: May 12, 2006, 03:29:41
Job time : 3260.76 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 19:15:19 ; Search time 142.787 Seconds
(without alignments)
12996.800 Million cell updates/sec

Title: US-09-423-093B-3_COPY_3977_5020
Perfect score: 1044
Sequence: 1 atgaatgtttttatcagtat.....ctccgcgtcttggtatttaa 1044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	56.2	5.4	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	2	52.6	5.0	1141	3	US-09-806-708B-22	Sequence 22, Appl
	3	48	4.6	50000	3	US-09-662-254B-24	Sequence 24, Appl
	4	47.2	4.5	601	3	US-09-949-016-115288	Sequence 115288,
c	5	47.2	4.5	5340	3	US-09-627-122-21	Sequence 21, Appl
	6	47.2	4.5	32392	3	US-09-662-254B-27	Sequence 27, Appl
c	7	47.2	4.5	50000	3	US-09-662-254B-23	Sequence 23, Appl
c	8	47	4.5	32392	3	US-09-662-254B-27	Sequence 27, Appl
	9	46.6	4.5	601	3	US-09-949-016-115287	Sequence 115287,
c	10	46.6	4.5	50000	3	US-09-662-254B-25	Sequence 25, Appl
	11	46.2	4.4	601	3	US-09-949-016-115290	Sequence 115290,
	12	46.2	4.4	601	3	US-09-949-016-115292	Sequence 115292,
	13	46.2	4.4	601	3	US-09-949-016-115294	Sequence 115294,
	14	46.2	4.4	601	3	US-09-949-016-115295	Sequence 115295,
	15	46.2	4.4	601	3	US-09-949-016-115296	Sequence 115296,
	16	46.2	4.4	135030	3	US-09-949-016-14896	Sequence 14896, A
c	17	46	4.4	1664976	3	US-08-916-421B-1	Sequence 1, Appli
c	18	46	4.4	1664976	3	US-09-692-570-1	Sequence 1, Appli
	19	45.8	4.4	601	3	US-09-949-016-115289	Sequence 115289,
	20	45.8	4.4	1830121	3	US-09-557-884-1	Sequence 1, Appli
	21	45.8	4.4	1830121	3	US-09-643-990A-1	Sequence 1, Appli
	22	45.8	4.4	1830121	3	US-10-158-865-1	Sequence 1, Appli
c	23	45.4	4.3	601	3	US-09-949-016-180812	Sequence 180812,
c	24	45.4	4.3	67620	3	US-09-949-016-16939	Sequence 16939, A
c	25	45.2	4.3	1410	3	US-09-662-254B-33	Sequence 33, Appl
c	26	45.2	4.3	50000	3	US-09-662-254B-24	Sequence 24, Appl
	27	45.2	4.3	319608	3	US-09-539-333D-1	Sequence 1, Appli
	28	45.2	4.3	319608	3	US-09-679-409-1	Sequence 1, Appli
	29	45	4.3	6124	3	US-08-213-419B-3	Sequence 3, Appli
	30	44.8	4.3	1752	3	US-09-710-279-1357	Sequence 1357, Ap
	31	44.8	4.3	3033	3	US-09-134-001C-2341	Sequence 2341, Ap
	32	44.8	4.3	3055	3	US-09-710-279-4075	Sequence 4075, Ap
	33	44.6	4.3	601	3	US-09-949-016-115291	Sequence 115291,
	34	44.6	4.3	601	3	US-09-949-016-115293	Sequence 115293,
c	35	44.6	4.3	3440	2	US-08-471-791-27	Sequence 27, Appl
c	36	44.6	4.3	3440	6	PCT-US91-01746-27	Sequence 27, Appl
	37	44.4	4.3	601	3	US-09-949-002-7549	Sequence 7549, Ap
	38	44.4	4.3	601	3	US-09-949-002-7550	Sequence 7550, Ap
	39	44.4	4.3	1353	3	US-09-248-796A-1080	Sequence 1080, Ap
c	40	44.4	4.3	2060	2	US-07-721-761A-31	Sequence 31, Appl
c	41	44.4	4.3	2060	2	US-07-978-687-31	Sequence 31, Appl
c	42	44.4	4.3	2060	6	PCT-US91-01750-2	Sequence 2, Appli
c	43	44.4	4.3	2060	6	PCT-US91-05801-31	Sequence 31, Appl
c	44	44.4	4.3	2081	6	PCT-US91-01750-3	Sequence 3, Appli
	45	44.4	4.3	3171	3	US-09-134-000C-1681	Sequence 1681, Ap

ALIGNMENTS

RESULT 1
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 5.4%; Score 56.2; DB 3; Length 1141;
Best Local Similarity 11.0%; Pred. No. 0.0013;
Matches 100; Conservative 332; Mismatches 473; Indels 7; Gaps 1;

```
Qy         92 GTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCCACAGAGAGATG 151
Db         58 KWTWARMYCKYRRWYNNKSRWWKGWYKKKWYBCANNTSBRYHARRWKDMKTAYBMTMTNK 117

Qy        152 AGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTA 211
Db        118 WGKTGWRHRYWRWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNATGWDD 177

Qy        212 ATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCATTAGTTTGACGA 271
Db        178 DTKYHMWNNNGCBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRK 237

Qy        272 CAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTAGCAGATGGAGCGTTATCAA 331
Db        238 VRRWVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRAKYWGW 297

Qy        332 AAATAGTGAAAGTTTTGAAGGCTAATCCTGAAATTGTATTGGCTACGCGAGCGTATGGTT 391
Db        298 NRABVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNWTY 357
```

```
Qy        392 GGTTTAAGGAAAATCCGAATGAGTTATGTGATACTGTTCGTCATTTAACAGACGATACTT 451
Db        358 KKARHBARWDWVWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKM 417

Qy        452 TATTTCAGCCGGGGGCTGATGCCATTAAATTTTTCTTCCGTAGAGTTGGAGTTATTTCAG 511
Db        418 DMDWBGTYNNNNNGGRTYYGWTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKT 477

Qy        512 GCTTTATTGTCAATGCTGAAAAAGCAAAAAAACTATCGAGTGATTTATTTGATGGGCGTT 571
Db        478 YWNNCYWKSMTNGKSHRBAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBVCSKWW 537

Qy        572 TATATTATCAAATGTACCT-------TGCTGGTATGCTAATGGCTGAAGGTCAGGGATAC 624
Db        538 NNYAAWYTKSSWNYTSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWS 597

Qy        625 TATTTTAGCGACGTGATGACATTGTCGAGGGATACAGAGGCTCCTGACTTTGGTAACGCT 684
Db        598 HTWBHBRAGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNNNNNATC 657

Qy        685 GGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTATAAACCAGAGGGCCGTATACATATG 744
Db        658 ARDDYYAASRWYAMANAKWYYYKBAANNAYYTHANNWWGCWNNATDTRRTMWKNNNNNNA 717

Qy        745 GTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAAGATACAACAAAAATTGATGGCGTT 804
Db        718 GTWKNNNNNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADAAABTTDKRNNGAY 777

Qy        805 TATGCTGGAATTAGAAAAGACTTAGCGAACTATTTTTATCCTTATATTCGAGATCAACTC 864
Db        778 TKYTTTNNNNTYRGVVTNTAARDGWANNNNNNNNNNNNNNGWSDMWVTWWAYANYGTNNN 837

Qy        865 GACTTGCCTCTTTATACTTATATTAAAATGATAAATAAATTTCGGAAAATGGGATTTTCA 924
Db        838 NNNNNNNAYAWWTNKWYYTTDDRWRBAYTNNNNNNRMAYYGAYADDYAYYMSDTCDAWMK 897

Qy        925 AATGAAAAGCTTTTCTATGTGCATGCCTTTTTAGGGTATGTACTAAAACGGAGGGGCTAT 984
Db        898 WDATKMNNATTYNRGTAWRTNNNNNNMTMKTKYYBHAAWNNNNNNGKMCTAHTWWVCKAT 957

Qy        985 GATGCTTTAATT 996
Db        958 KTTKGCWMNCTT 969
```

RESULT 2
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 5.0%; Score 52.6; DB 3; Length 1141;
Best Local Similarity 13.4%; Pred. No. 0.0087;
Matches 109; Conservative 289; Mismatches 401; Indels 15; Gaps 4;

```
Qy         79 AATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCT 138
Db       1138 WRYTCHGNTTTKTKTYKKANNNNNNNNGMGKDWNRMDATKWSATGTAWWTNHAKRGATMC 1079

Qy        139 CCACAGAGAGATGAGATAAACTCTATT-----ATCGAAAACTATAAAGCAAAAAATAATA 193
Db       1078 WYWYWTGTNRRWCMRTYAMRTWYTRSNANWSCATKBMWWTMKWYATKYRTAWYAMWCAWR 1019

Qy        194 AACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAA 253
Db       1018 NNNMWCATNGYAKSCATNNAMWYATTRWAAYAAAKWARWAGNNMRMYGAAAGNKWGCMAA 959

Qy        254 AATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTAG 313
Db        958 MATMGBWWADTAGKMC----NNNNNNWTTDVRRMAMKAKNNNNNNAYWTACYNRAATNNK 903

Qy        314 CAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAAATTGTATTGG 373
Db        902 MATHWMKWTHGAHSKRRTRHHTRTCRRTKYNNNNNNARTVYWYHHAARRWMNAWWTRTNN 843

Qy        374 CTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGTGATACTGTTCGTC 433
Db        842 NNNNNNNNACRNTRTWWABWKHSWCNNNNNNNNNNNNNNNTWCHYTTANABBCYRANNNNA 783

Qy        434 ATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAATTTTTCTTCCGTA 493
Db        782 AARMARTCNNYMHAAVTTTHTDWCYKTWMNTWYWDMMTTMBTTTTRNMTTSTNMTNNNNN 723

Qy        494 GAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAAAAACTATCGAGTG 553
Db        722 NMWA---CTNNNNNNMWKAYYAHATNNWGCWWNNTDARRTNNTTVMRRRWMTNTKTRWYS 666

Qy        554 ATTTATTTGATGGGCGTTTATATTATCAAATGTACCTTGCTGGTATGCTAATGGCTGAAG 613
Db        665 TTRRHHYTGATNNNNNNNNNNNNNNNNSCCTCTRMMTMRWTMKGDGMTVRKKVKWRDTTC 606

Qy        614 GTCAGGGATACTATTTTAGCGACGTGATGACAT---TGTCGAGGGATACAGAGGCTCCTG 670
Db        605 TYVDVWADSWVWWYANWMRCRDVTYTRNNTYCKSYAHSYWYWSNNAMWYRRYSARNWSSM 546

Qy        671 ACTTTGGTAACGCTGGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTATAAACCAGAGG 730
Db        545 ARWTTRNNWWMSGBVRMRWAGTMWWRHWNNNNTDTRYYWWWKRWARBTTTVYDSMCNAKS 486

Qy        731 GCCGTATACATATGGTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAAGATACAACAA 790
Db        485 MWRGNNWRAMKMWWAANNDAGAMDHWTYWMGNNTMWMRRAWKMMNMAWCRRAYCCNNNNN 426

Qy        791 AAATTGATGGCGTTTATGCTGGAATTAGAAAAGACTTAGCGAACTATTTTTATCCTTATA 850
Db        425 RACVWHKHKMWRWTWKYMWKAACNNNNBKAMYMRVAWMMYSRDTTNTDWMMWTSDWBWHW 366

Qy        851 TTCGAGATCAACTCGACTTGCCTCTTTATACTTA 884
Db        365 YTVDYTMMRAWNNNNNNNWRBCKTTSWMWWMDHM 332
```

RESULT 3
US-09-662-254B-24
; Sequence 24, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29

```
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-24

  Query Match             4.6%;  Score 48;  DB 3;  Length 50000;
  Best Local Similarity   50.4%;  Pred. No. 0.22;
  Matches  117;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy         33 TTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATATATAATCAAGATTATTG 92
              || ||||  |  | |   |  ||||   |||       | |   | | |  ||| |  ||
Db      19952 TTTTAATCAAATTAATAATACAGAGTTTCTAGAAATATGTAATAAAAGTTTAGAATCATG 20011

Qy         93 TTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCCACAGAGAGATGA 152
              ||||  |||  |||| | |     || ||| | |  || |   |     |         |
Db      20012 GATAAAACATATAGATTGTAAAAAATTTATTTTTAATAATGATAGTTATTATGATTTGCA 20071

Qy        153 GATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAA 212
              |||  |  ||| |  || |  ||  |    | ||||||||  | | | |   |  | |
Db      20072 AATATCCGTTATAAATGATATGAATGTATTTGATAATAATAATAATATTATAGCTTATTA 20131

Qy        213 TTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCATTAGT 264
              ||  ||| || ||| |  |    || |||| ||| | ||||||   |||| |
Db      20132 TTATAATAAACATATTAAAATAAATTATAATAATATTAAAAAACATATTAAT 20183


RESULT 4
US-09-949-016-115288
; Sequence 115288, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 115288
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-115288

  Query Match             4.5%;  Score 47.2;  DB 3;  Length 601;
  Best Local Similarity   48.1%;  Pred. No. 0.13;
  Matches  130;  Conservative    1;  Mismatches  139;  Indels    0;  Gaps    0;

Qy         13 ATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGC 72
              |: | ||| | |||| | | |||| ||| |  | |  | |    |   | |  |||
Db        300 AWAAATATATATATAACATATTATTATATATATTATATATATTACTATATGTTATATATA 359

Qy         73 ATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGAT 132
              |||||| ||  | |||||   ||  | | ||     |   |  |   | ||     |  ||
Db        360 ATATATTATATATATTATATATTCTATATTATAATATAGAATATATATATTCTATATTAT 419

Qy        133 AAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAAT 192
              ||       | | | |    | ||||  | ||||||  | |   || |   |||  |||
Db        420 AATATATTATATATAATATATATAATATATTATATATATAGATTATATAAATTATATG 479

Qy        193 AAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAA 252
              || | |   | ||   ||  ||  || ||    || ||   |||  ||  |  |  |
Db        480 TAATATATATTATAATATAGATTATATAAATTATATGTAATATATATTATAATATAGATT 539

Qy        253 AAATGCATTAGTTTGACGACAGGTAAATAT 282
              | ||  ||||  |  |  | |  | |  ||
Db        540 ATATAAATTATATGTAATATATATTATAAT 569


RESULT 5
US-09-627-122-21/c
; Sequence 21, Application US/09627122
; Patent No. 6472521
; GENERAL INFORMATION:
;  APPLICANT:  Uhlmann, Eugen
;  APPLICANT:  Greiner, Beate
;  APPLICANT:  Unger, Eberhard
;  APPLICANT:  Gothe, Gislinde
;  APPLICANT:  Schwerdel, Marc
;  TITLE OF INVENTION:  OLIGONUCLEOTIDES FOR THE INHIBITION OF HUMAN eg5
;  TITLE OF INVENTION:  EXPRESSION
;  FILE REFERENCE:  02481.1678
;  CURRENT APPLICATION NUMBER: US/09/627,122
;  CURRENT FILING DATE:  2000-07-27
;  NUMBER OF SEQ ID NOS:  23
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 21
;   LENGTH: 5340
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
US-09-627-122-21

  Query Match             4.5%;  Score 47.2;  DB 3;  Length 5340;
  Best Local Similarity   49.7%;  Pred. No. 0.21;
  Matches  150;  Conservative    0;  Mismatches  148;  Indels    4;  Gaps    1;

Qy          1 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 60
              ||  || | ||||      ||    ||  | ||    |||| | |  |  ||  ||  |
Db       1178 ATAGATATGTTTACTCCGATGAAAATCACATCACTAAATAGGGATTCGACTTACGACACG 1119

Qy         61 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 120
                ||||  |   | ||| | ||| ||  |  ||| ||| |||||  | |    | |
Db       1118 AATTATGATAATAATAAAAATGATGATAATAATAAAAATAATCATGATAATAATAAAAAT 1059

Qy        121 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 180
                |  |||  ||||         |    ||| | |  ||   |  |  | | ||  | ||
Db       1058 AATCATGATAATAATAAAAATAATCATGATAATAATAAAAATGGTGACAATAATAAAAAT 999

Qy        181 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 240
              |   | | |||||  |||||  | |    ||||     |||| |||||  |    || |
Db        998 GATCATGACAATAATAAAAATGGTGACAATAATAATTATGATGATAATAAAATGGATCAA 939

Qy        241 ----AAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGG 296
                | ||   ||| ||    | |||   ||| || |   ||||    || | |||||
Db        938 GGTCATGATCATAATAATTATCTTTATAATGATGATAATAAAATGGATCAAGGTCATGAT 879

Qy        297 CA 298
              ||
Db        878 CA 877


RESULT 6
US-09-662-254B-27
; Sequence 27, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
```

```
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
;   LENGTH: 32392
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-27

  Query Match             4.5%;  Score 47.2;  DB 3;  Length 32392;
  Best Local Similarity   52.6%;  Pred. No. 0.31;
  Matches  103;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy         67 GATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGT 126
              ||||  |||  ||||    |||||  | ||||| |||  ||||  |
Db      29030 GATAATATAGTTAATGTTTATTATAGATTAAATAATTGTGATAATATCAGATCAAGAGAA 29089

Qy        127 GAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAA 186
              | | ||   |    | |   |||| |   |   ||||||     |  |||  | ||||
Db      29090 ATACAAAATAGTTTTAAAATGTATGAAAATTATAATATTATGTTTTATAATAGTGAAAAA 29149

Qy        187 AATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAAT 246
              ||| ||||   | |  | |||  |||||  ||| |  ||||||| |||| |    | ||
Db      29150 AATTATAACATAGACATATATTGTAATTATAATAATAATAATTTCGGCTTTATGTATAAA 29209

Qy        247 TTAAAAAAATGCATTA 262
              ||| ||    || |
Db      29210 AAAAATAATAAAATAA 29225


RESULT 7
US-09-662-254B-23/c
; Sequence 23, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-23

  Query Match             4.5%;  Score 47.2;  DB 3;  Length 50000;
  Best Local Similarity   52.6%;  Pred. No. 0.34;
  Matches  103;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy         67 GATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGT 126
              ||||  |||  ||||    |||||  | ||||| |||  ||||  |
Db       3363 GATAATATAGTTAATGTTTATTATAGATTAAATAATTGTGATAATATCAGATCAAGAGAA 3304

Qy        127 GAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAA 186
              | | ||   |    | |   |||| |   |   ||||||     |  |||  | ||||
Db       3303 ATACAAAATAGTTTTAAAATGTATGAAAATTATAATATTATGTTTTATAATAGTGAAAAA 3244

Qy        187 AATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAAT 246
              ||| ||||   | |  | |||  |||||  ||| |  ||||||| |||| |    | ||
Db       3243 AATTATAACATAGACATATATTGTAATTATAATAATAATAATTTCGGCTTTATGTATAAA 3184

Qy        247 TTAAAAAAATGCATTA 262
              ||| ||    || |
Db       3183 AAAAATAATAAAATAA 3168


RESULT 8
US-09-662-254B-27/c
; Sequence 27, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
;   LENGTH: 32392
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-27

  Query Match             4.5%;  Score 47;  DB 3;  Length 32392;
  Best Local Similarity   53.2%;  Pred. No. 0.34;
  Matches  126;  Conservative    0;  Mismatches  105;  Indels    6;  Gaps    1;

Qy         11 TTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATA 70
              || | | ||    |||   | |||||| ||||| ||  || |||||   | |     |||
Db      16260 TTTTTAATACAAATATTATGACTTATAGTAGAGATGTATTATTAGAAATAATCAATAATA 16201

Qy         71 GCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAG 130
              ||||| | | | ||||   | | |  ||| | ||  |     ||  ||    ||
Db      16200 ------ATAATTACGTTGATTGCGAACATAGAAATAACTTATATAATATATTTAAACAAA 16147

Qy        131 ATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATA 190
              | ||   | |  |   | ||   |  || | ||||    |||   ||||  | ||||||
Db      16146 AAAATCATTCTAATGAAAATATTAATAATTTTATTTATAAAATAGATAATACGAAAAATT 16087

Qy        191 ATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |  | | |||  | |||||  |    || |  ||| | ||    ||||||| || |
Db      16086 TTTCAAATAATAATAATGTTTTTAAATATAATAATAGTATAATACATGATAAAAAAT 16030


RESULT 9
US-09-949-016-115287
; Sequence 115287, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 115287
;   LENGTH: 601
```

```
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-115287

  Query Match             4.5%;  Score 46.6;  DB 3;  Length 601;
  Best Local Similarity   48.0%;  Pred. No. 0.18;
  Matches  130;  Conservative    1;  Mismatches  140;  Indels    0;  Gaps    0;

Qy         12 TATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAG 71
              :|  | ||| | |||| | | |||| ||| |  | |  | |    |      |  |  |||
Db        301 WAAAAATATATATATAACATATTATTATATATATTATATATATTACTATATGTTATATAT 360

Qy         72 CATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGA 131
              |||||| ||  | |||||   ||  | | ||      |   |  |   | ||    |  |
Db        361 AATATATTATATATATTATATATTCTATATTATAATATAGAATATATATATTCTATATTA 420

Qy        132 TAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAA 191
              |||       | | | |    | ||||  | ||||||  | |   || |   |||  |||
Db        421 TAATATATTATATATAATATATAATATATATTATATATATAGATTATATAAATTATAT 480

Qy        192 TAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAA 251
              || | |   | ||   ||  ||  || ||    || ||   |||  ||  |  |  |
Db        481 GTAATATATATTATAATATAGATTATATAAATTATATGTAATATATATTATAATATAGAT 540

Qy        252 AAAATGCATTAGTTTGACGACAGGTAAATAT 282
              | ||  |||| |  |  | |  | |  ||
Db        541 TATATAAATTATATGTAATATATATTATAAT 571


RESULT 10
US-09-662-254B-25/c
; Sequence 25, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-25

  Query Match             4.5%;  Score 46.6;  DB 3;  Length 50000;
  Best Local Similarity   50.7%;  Pred. No. 0.47;
  Matches  112;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy         62 TACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTG 121
              || |  |||  | | ||    || || ||  |||||| ||||  |  || ||  |    |
Db      39801 TATTACATATGAAAGATGGAGAAAATAATATTTTAAAAAATAGAGTATTCGAATTACCAG 39742

Qy        122 TTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAG 181
                |  | ||||          |      | |||||  |  | ||| |||||   |  |
Db      39741 AAGTTATAAATAATGTATATATTAATAGATATATAAATTTAAATATTGAAAATATTGTAT 39682

Qy        182 CAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATA 241
              | ||  || || | |||||  |  ||  |||| | |||| | |  ||||    |||  |
Db      39681 TAGAATCTATTAGAAAAAATTTCAATAATAATATAAATGTACAAGATTTGATATATAGTT 39622

Qy        242 AGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATAT 282
              |  | | |  ||    ||| | | |  | |    || |||
Db      39621 TAAGATCAGATCAAGAATTTAATATAAATATAAAAAATTAT 39581


RESULT 11
US-09-949-016-115290
; Sequence 115290, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 115290
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-115290

  Query Match             4.4%;  Score 46.2;  DB 3;  Length 601;
  Best Local Similarity   48.3%;  Pred. No. 0.22;
  Matches  129;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy         16 AGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATA 75
              | ||| | |||| | | |||| ||| |  | |  | |    |      |  |  |||  |||
Db        163 AATATATATATAACATATTATTATATATATTATATATATTACTATATGTTATATATAATA 222

Qy         76 TATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAA 135
              ||| ||  | |||||   ||  | | ||      |  |  |   | ||    |  ||||
Db        223 TATTATATATATTATATATTCTATATTATAATATAGAATATATATATTCTATATTATAAT 282

Qy        136 TCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAA 195
                | | | |    | ||||  | ||||||  | |   || |   |||  |||   ||
Db        283 ATATTATATATAATATATATAATATATATTATATATATAGATTATATAAATTATATGTAA 342

Qy        196 CAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAA 255
              | |   | ||   ||  ||  || ||    || ||   |||  ||  |  |  |  | |
Db        343 TATATATTATAATATAGATTATATAAATTATATGTAATATATATTATAATATAGATTATA 402

Qy        256 TGCATTAGTTTGACGACAGGTAAATAT 282
              |  ||||  |  |  | |  | |  ||
Db        403 TAAATTATATGTAATATATATTATAAT 429


RESULT 12
US-09-949-016-115292
; Sequence 115292, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 115292
;   LENGTH: 601
```

```
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-115292

  Query Match             4.4%;  Score 46.2;  DB 3;  Length 601;
  Best Local Similarity   48.3%;  Pred. No. 0.22;
  Matches  129;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy         16 AGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATA 75
              | ||| | |||| | | |||| ||| |  | |  | |    |     |  |  |||  |||
Db        156 AATATATATATAACATATTATTATATATATTATATATATTACTATATGTTATATATAATA 215

Qy         76 TATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAA 135
              ||| ||  | |||||   ||  | | ||      |   |  |   | ||     |  ||||
Db        216 TATTATATATATTATATATTCTATATTATAATATAGAATATATATATTCTATATTATAAT 275

Qy        136 TCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAA 195
              | | | |   | ||||  | ||||||  | |   || |   |||  |||   ||
Db        276 ATATTATATATAATATATATAATATATATTATATATATAGATTATATAAATTATATGTAA 335

Qy        196 CAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAA 255
              | |   | ||   ||  ||  || ||    || ||   |||  ||  |  |  |  | |
Db        336 TATATATTATAATATAGATTATATAAATTATATGTAATATATATTATAATATAGATTATA 395

Qy        256 TGCATTAGTTTGACGACAGGTAAATAT 282
              |  ||||  |  |  | |  | |  ||
Db        396 TAAATTATATGTAATATATATTATAAT 422


RESULT 13
US-09-949-016-115294
; Sequence 115294, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 115294
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-115294

  Query Match             4.4%;  Score 46.2;  DB 3;  Length 601;
  Best Local Similarity   48.3%;  Pred. No. 0.22;
  Matches  129;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy         16 AGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATA 75
              | ||| | |||| | | |||| ||| |  | |  | |    |     |  |  |||  |||
Db        150 AATATATATATAACATATTATTATATATATTATATATATTACTATATGTTATATATAATA 209

Qy         76 TATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAA 135
              ||| ||  | |||||   ||  | | ||      |   |  |   | ||     |  ||||
Db        210 TATTATATATATTATATATTCTATATTATAATATAGAATATATATATTCTATATTATAAT 269

Qy        136 TCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAA 195
              | | | |   | ||||  | ||||||  | |   || |   |||  |||   ||
Db        270 ATATTATATATAATATATATAATATATATTATATATATAGATTATATAAATTATATGTAA 329

Qy        196 CAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAA 255
              | |   | ||   ||  ||  || ||    || ||   |||  ||  |  |  |  | |
Db        330 TATATATTATAATATAGATTATATAAATTATATGTAATATATATTATAATATAGATTATA 389

Qy        256 TGCATTAGTTTGACGACAGGTAAATAT 282
              |  ||||  |  |  | |  | |  ||
Db        390 TAAATTATATGTAATATATATTATAAT 416


RESULT 14
US-09-949-016-115295
; Sequence 115295, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 115295
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-115295

  Query Match             4.4%;  Score 46.2;  DB 3;  Length 601;
  Best Local Similarity   48.3%;  Pred. No. 0.22;
  Matches  129;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy         16 AGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATA 75
              | ||| | |||| | | |||| ||| |  | |  | |    |     |  |  |||  |||
Db        148 AATATATATATAACATATTATTATATATATTATATATATTACTATATGTTATATATAATA 207

Qy         76 TATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAA 135
              ||| ||  | |||||   ||  | | ||      |   |  |   | ||     |  ||||
Db        208 TATTATATATATTATATATTCTATATTATAATATAGAATATATATATTCTATATTATAAT 267

Qy        136 TCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAA 195
              | | | |   | ||||  | ||||||  | |   || |   |||  |||   ||
Db        268 ATATTATATATAATATATATAATATATATTATATATATAGATTATATAAATTATATGTAA 327

Qy        196 CAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAA 255
              | |   | ||   ||  ||  || ||    || ||   |||  ||  |  |  |  | |
Db        328 TATATATTATAATATAGATTATATAAATTATATGTAATATATATTATAATATAGATTATA 387

Qy        256 TGCATTAGTTTGACGACAGGTAAATAT 282
              |  ||||  |  |  | |  | |  ||
Db        388 TAAATTATATGTAATATATATTATAAT 414


RESULT 15
US-09-949-016-115296
; Sequence 115296, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
```

```
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 115296
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-115296

  Query Match             4.4%;  Score 46.2;  DB 3;  Length 601;
  Best Local Similarity   48.3%;  Pred. No. 0.22;
  Matches  129;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy         16 AGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATA 75
              | ||| | |||| | | |||| ||| |  | |  | |   |      |  |  |||  |||
Db        147 AATATATATATAACATATTATTATATATATTATATATATTACTATATGTTATATATAATA 206

Qy         76 TATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAA 135
              ||| ||  | |||||   ||  | | ||      |   |  |   | ||    |  ||||
Db        207 TATTATATATATTATATATTCTATATTATAATATAGAATATATATATTCTATATTATAAT 266

Qy        136 TCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAA 195
                | | | |    | ||||  | ||||||  | |   || |   |||  |||   ||
Db        267 ATATTATATATAATATATATAATATATATTATATATATAGATTATATAAATTATATGTAA 326

Qy        196 CAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAA 255
               | |   | ||   ||  ||  || ||    || ||   |||  ||  |  |  |  | |
Db        327 TATATATTATAATATAGATTATATAAATTATATGTAATATATATTATAATATAGATTATA 386

Qy        256 TGCATTAGTTTGACGACAGGTAAATAT 282
              |  ||||  |  |  | |  | |  ||
Db        387 TAAATTATATGTAATATATATTATAAT 413


Search completed: May 12, 2006, 03:51:19
Job time : 148.787 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:31:18 ; Search time 668.458 Seconds
(without alignments)
12915.142 Million cell updates/sec

Title: US-09-423-093B-3_COPY_3977_5020
Perfect score: 1044
Sequence: 1 atgaatgtttttatcagtat.....ctccgcgtcttggtatttaa 1044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1044	100.0	12441	3	US-09-423-093A-3	Sequence 3, Appli
	2	55.8	5.3	1404	7	US-10-282-122A-16075	Sequence 16075, A
	3	55.6	5.3	564	7	US-10-437-963-7175	Sequence 7175, Ap
c	4	53.6	5.1	6642	10	US-11-097-143-5146	Sequence 5146, Ap
c	5	53.6	5.1	10992	10	US-11-097-143-5104	Sequence 5104, Ap
c	6	52.8	5.1	3089	9	US-10-954-778-160	Sequence 160, App
c	7	52.8	5.1	4985	5	US-10-094-240-10	Sequence 10, Appl
c	8	52.8	5.1	4985	6	US-10-056-405-10	Sequence 10, Appl
	9	51.6	4.9	11143	10	US-11-097-143-16492	Sequence 16492, A
	10	51	4.9	441	7	US-10-437-963-98795	Sequence 98795, A
	11	50.8	4.9	337344	8	US-10-388-838-58	Sequence 58, Appl
	12	48.8	4.7	4042	8	US-10-775-920-306	Sequence 306, App
c	13	48.8	4.7	6179	7	US-10-221-714A-66	Sequence 66, Appl
	14	48.4	4.6	1127	5	US-10-027-632-119520	Sequence 119520,
	15	48.4	4.6	1127	6	US-10-027-632-119520	Sequence 119520,
	16	48.2	4.6	8801	5	US-10-239-676-144	Sequence 144, App
	17	48.2	4.6	8801	6	US-10-311-455-1714	Sequence 1714, Ap
	18	48.2	4.6	8801	6	US-10-240-453-160	Sequence 160, App
c	19	48.2	4.6	9963	6	US-10-311-455-668	Sequence 668, App
	20	48	4.6	50000	8	US-10-706-635-24	Sequence 24, Appl
	21	47.8	4.6	405	7	US-10-437-963-13671	Sequence 13671, A
	22	47.6	4.6	5376	6	US-10-311-455-2123	Sequence 2123, Ap
c	23	47.6	4.6	11812	5	US-10-239-676-209	Sequence 209, App
c	24	47.6	4.6	11812	6	US-10-311-455-2091	Sequence 2091, Ap
c	25	47.6	4.6	11812	6	US-10-240-453-305	Sequence 305, App
c	26	47.6	4.6	11812	7	US-10-221-714A-467	Sequence 467, App
	27	47.4	4.5	7108	7	US-10-257-166-77	Sequence 77, Appl
	28	47.4	4.5	18624	6	US-10-311-455-1675	Sequence 1675, Ap
	29	47.2	4.5	32392	8	US-10-706-635-27	Sequence 27, Appl
c	30	47.2	4.5	50000	8	US-10-706-635-23	Sequence 23, Appl
c	31	47.2	4.5	3673778	6	US-10-312-841-2	Sequence 2, Appli
c	32	47	4.5	8413	6	US-10-240-485-50	Sequence 50, Appl
c	33	47	4.5	32392	8	US-10-706-635-27	Sequence 27, Appl
c	34	46.8	4.5	7442	7	US-10-221-714A-409	Sequence 409, App
	35	46.6	4.5	37515	7	US-10-433-793-28	Sequence 28, Appl
c	36	46.6	4.5	50000	8	US-10-706-635-25	Sequence 25, Appl
c	37	46.4	4.4	264	3	US-09-969-373-472	Sequence 472, App
c	38	46.4	4.4	128034	6	US-10-282-174-186	Sequence 186, App
c	39	46.4	4.4	128034	6	US-10-282-174-187	Sequence 187, App
c	40	46.4	4.4	128034	8	US-10-600-009-186	Sequence 186, App
c	41	46.4	4.4	128034	8	US-10-600-009-187	Sequence 187, App
	42	46.4	4.4	202100	6	US-10-282-174-484	Sequence 484, App
	43	46.4	4.4	202100	8	US-10-600-009-484	Sequence 484, App
	44	46	4.4	5397	6	US-10-311-455-1655	Sequence 1655, Ap
c	45	46	4.4	6486	7	US-10-433-793-80	Sequence 80, Appl

ALIGNMENTS

RESULT 1
US-09-423-093A-3
; Sequence 3, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12441
; TYPE: DNA
; ORGANISM: Salmonella enterica
US-09-423-093A-3

Query Match 100.0%; Score 1044; DB 3; Length 12441;
Best Local Similarity 100.0%; Pred. No. 2.1e-207;
Matches 1044; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3977 ATGAATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCA 4036

Qy         61 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4037 CTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATT 4096

Qy        121 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4097 GTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAA 4156

Qy        181 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4157 GCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGAT 4216

Qy        241 AAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4217 AAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAAC 4276
```

```
Qy        301 GATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4277 GATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCT 4336

Qy        361 GAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4337 GAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGT 4396

Qy        421 GATACTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4397 GATACTGTTCGTCATTTAACAGACGATACTTTATTTCAGCCGGGGGCTGATGCCATTAAA 4456

Qy        481 TTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4457 TTTTTCTTCCGTAGAGTTGGAGTTATTTCAGGCTTTATTGTCAATGCTGAAAAAGCAAAA 4516

Qy        541 AAACTATCGAGTGATTTATTTGATGGGCGTTTATATTATCAAATGTACCTTGCTGGTATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4517 AAACTATCGAGTGATTTATTTGATGGGCGTTTATATTATCAAATGTACCTTGCTGGTATG 4576

Qy        601 CTAATGGCTGAAGGTCAGGGATACTATTTTAGCGACGTGATGACATTGTCGAGGGATACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4577 CTAATGGCTGAAGGTCAGGGATACTATTTTAGCGACGTGATGACATTGTCGAGGGATACA 4636

Qy        661 GAGGCTCCTGACTTTGGTAACGCTGGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4637 GAGGCTCCTGACTTTGGTAACGCTGGAACTGAAAAAGGAGTTTTCACCCCGGGGGGGTAT 4696

Qy        721 AAACCAGAGGGCCGTATACATATGGTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4697 AAACCAGAGGGCCGTATACATATGGTTGAAGGCTTGTTGCTAATTGCAAAATATATAGAA 4756

Qy        781 GATACAACAAAAATTGATGGCGTTTATGCTGGAATTAGAAAAGACTTAGCGAACTATTTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4757 GATACAACAAAAATTGATGGCGTTTATGCTGGAATTAGAAAAGACTTAGCGAACTATTTT 4816

Qy        841 TATCCTTATATTCGAGATCAACTCGACTTGCCTCTTTATACTTATATTAAAATGATAAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4817 TATCCTTATATTCGAGATCAACTCGACTTGCCTCTTTATACTTATATTAAAATGATAAAT 4876

Qy        901 AAATTTCGGAAAATGGGATTTTCAAATGAAAAGCTTTTCTATGTGCATGCCTTTTTAGGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4877 AAATTTCGGAAAATGGGATTTTCAAATGAAAAGCTTTTCTATGTGCATGCCTTTTTAGGG 4936

Qy        961 TATGTACTAAAACGGAGGGGCTATGATGCTTTAATTAAATACATTCGTAGCAAAAAAGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4937 TATGTACTAAAACGGAGGGGCTATGATGCTTTAATTAAATACATTCGTAGCAAAAAAGGC 4996

Qy       1021 GGTACTCCGCGTCTTGGTATTTAA 1044
              ||||||||||||||||||||||||
Db       4997 GGTACTCCGCGTCTTGGTATTTAA 5020
```

RESULT 2
US-10-282-122A-16075
; Sequence 16075, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16075
; LENGTH: 1404
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-282-122A-16075

Query Match 5.3%; Score 55.8; DB 7; Length 1404;
Best Local Similarity 54.0%; Pred. No. 0.19;
Matches 114; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```
Qy         74 TATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATA 133
              |||||||  ||||| ||| | |||||||     ||  ||       ||    | ||
Db        965 TATATAAAGAAGATAATTTTATAAAGAAGGTAAATAATGTTCAGTGTGAACTTAAAATAG 1024

Qy        134 AATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATA 193
              || |   ||   ||| || ||||   || | |  |||  |||    |||| || |||
Db       1025 GATTTAATCAAGTAGAAGATTTAAATATTAATTTAAAAATATATTCTATAAAAGATCATA 1084

Qy        194 AACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAA 253
              |  | ||| |  ||| ||||    ||  | |  |||| |  |||||||       ||
Db       1085 ATTATAATATAAATGATAATAAATATATACAGTATTTTGATTATGATAAAGTAAGTAATG 1144

Qy        254 AATGCATTAGTTTGACGACAGGTAAATATTG 284
              ||   |||  ||| | |   | ||| || |
Db       1145 AAAAAATTTATTTAAGAAATAGAAAAAATGG 1175
```

RESULT 3
US-10-437-963-7175
; Sequence 7175, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 7175
; LENGTH: 564

```
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_13795C.1
US-10-437-963-7175

  Query Match             5.3%;  Score 55.6;  DB 7;  Length 564;
  Best Local Similarity   52.1%;  Pred. No. 0.14;
  Matches  124;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||  |   ||||| |  || ||  |   || |||||| ||  | |    | |  |  |
Db        268 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 327

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |   || | |  ||   || ||   | ||  ||||    || |
Db        328 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 387

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| |  |||||       || |||| |||
Db        388 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 447

Qy        248 TAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGA 305
                | ||     |||  | | ||  | | ||  |||  ||| |    |   |||| || |
Db        448 ATAGAAGTACTATTGATATCACCTCGGCTAGTTATGTCATCAAATTTTCCAACAATAA 505


RESULT 4
US-11-097-143-5146/c
; Sequence 5146, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5146
;   LENGTH: 6642
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-5146

  Query Match             5.1%;  Score 53.6;  DB 10;  Length 6642;
  Best Local Similarity   55.3%;  Pred. No. 1;
  Matches  104;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              ||| ||   ||||| |  || ||  |   || |||||| ||  | |    | |  |  |
Db       6632 ATACCAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 6573

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |   || | |  ||   || ||   | ||  ||||    || |
Db       6572 ATAATAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAACAATA 6513

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| ||  |||||       || |||||||
Db       6512 ATAATAATAATAATAATAATAATAATAATAATAAAATTAATTATAATGATAATAAGAACT 6453

Qy        248 TAAAAAAA 255
              ||| || |
Db       6452 TAACAACA 6445


RESULT 5
US-11-097-143-5104/c
; Sequence 5104, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5104
;   LENGTH: 10992
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-5104

  Query Match             5.1%;  Score 53.6;  DB 10;  Length 10992;
  Best Local Similarity   55.3%;  Pred. No. 1.3;
  Matches  104;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              ||| ||   ||||| |  || ||  |   || |||||| ||  | |    | |  |  |
Db      10982 ATACCAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 10923

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |   || | |  ||   || ||   | ||  ||||    || |
Db      10922 ATAATAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAACAATA 10863

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| ||  |||||       || |||||||
Db      10862 ATAATAATAATAATAATAATAATAATAATAATAAAATTAATTATAATGATAATAAGAACT 10803

Qy        248 TAAAAAAA 255
              ||| || |
Db      10802 TAACAACA 10795


RESULT 6
US-10-954-778-160/c
; Sequence 160, Application US/10954778
```

```
; Publication No. US20050153368A1
; GENERAL INFORMATION:
;  APPLICANT: ZWIEBEL, LAURENCE J.
;  TITLE OF INVENTION: METHOD OF IDENTIFYING CHEMICAL AGENTS WHICH STIMULATE
;  TITLE OF INVENTION:  ODORANT RECEPTORS OF SENSORY NEURONS
;  FILE REFERENCE: N1125
;  CURRENT APPLICATION NUMBER: US/10/954,778
;  CURRENT FILING DATE:  2004-09-30
;  PRIOR APPLICATION NUMBER: 10/056,405
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: 60/264,649
;  PRIOR FILING DATE: 2001-01-26
;  NUMBER OF SEQ ID NOS: 248
;  SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 160
;   LENGTH: 3089
;   TYPE: DNA
;   ORGANISM: Anopheles gambiae
US-10-954-778-160

  Query Match             5.1%;  Score 52.8;  DB 9;  Length 3089;
  Best Local Similarity   54.0%;  Pred. No. 1.1;
  Matches  108;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||  |   |||||  |  || ||  |   || ||||||  ||  | |    | | ||  |
Db       2050 ATAAAAGTAATAATTATAATAATAATAGTAATAATAATAATAATAATAATAATTAATA 1991

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |   || | |  ||   || |    | ||  ||||    || |
Db       1990 ATAATAATAATAATAATAATAATAATAATAATAATAATTATAATAATAATAATAATAATA 1931

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| |  |||||       || |||| |||
Db       1930 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1871

Qy        248 TAAAAAAATGCATTAGTTTG 267
                || |||| |  || ||||
Db       1870 ATAATAAATTCTGTATTTTG 1851


RESULT 7
US-10-094-240-10/c
; Sequence 10, Application US/10094240
; Publication No. US20030082637A1
; GENERAL INFORMATION:
;  APPLICANT: ZWIEBEL, LAURENCE J.
;  TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
;  FILE REFERENCE: N8289
;  CURRENT APPLICATION NUMBER: US/10/094,240
;  CURRENT FILING DATE:  2001-03-08
;  PRIOR APPLICATION NUMBER: 10/056,405
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: 60/264,649
;  PRIOR FILING DATE: 2001-01-26
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 4985
;   TYPE: DNA
;   ORGANISM: Anopheles gambiae
US-10-094-240-10

  Query Match             5.1%;  Score 52.8;  DB 5;  Length 4985;
  Best Local Similarity   54.0%;  Pred. No. 1.4;
  Matches  108;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||  |   |||||  |  || ||  |   || ||||||  ||  | |    | | ||  |
Db       2535 ATAAAAGTAATAATTATAATAATAATAGTAATAATAATAATAATAATAATAATTAATA 2476

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |   || | |  ||   || |    | ||  ||||    || |
Db       2475 ATAATAATAATAATAATAATAATAATAATAATAATAATTATAATAATAATAATAATAATA 2416

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| |  |||||       || |||| |||
Db       2415 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2356

Qy        248 TAAAAAAATGCATTAGTTTG 267
                || |||| |  || ||||
Db       2355 ATAATAAATTCTGTATTTTG 2336


RESULT 8
US-10-056-405-10/c
; Sequence 10, Application US/10056405
; Publication No. US20030166013A1
; GENERAL INFORMATION:
;  APPLICANT: ZWIEBEL, LAURENCE J.
;  TITLE OF INVENTION: MOSQUITO OLFACTORY GENES, POLYPEPTIDES, AND METHODS OF
;  TITLE OF INVENTION:  USE THEREOF
;  FILE REFERENCE: N7841
;  CURRENT APPLICATION NUMBER: US/10/056,405
;  CURRENT FILING DATE:  2002-01-24
;  PRIOR APPLICATION NUMBER: 60/264,649
;  PRIOR FILING DATE: 2001-01-26
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 4985
;   TYPE: DNA
;   ORGANISM: Anopheles gambiae
US-10-056-405-10

  Query Match             5.1%;  Score 52.8;  DB 6;  Length 4985;
  Best Local Similarity   54.0%;  Pred. No. 1.4;
  Matches  108;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||  |   |||||  |  || ||  |   || ||||||  ||  | |    | | ||  |
Db       2535 ATAAAAGTAATAATTATAATAATAATAGTAATAATAATAATAATAATAATAATTAATA 2476

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |   || | |  ||   || |    | ||  ||||    || |
Db       2475 ATAATAATAATAATAATAATAATAATAATAATAATAATTATAATAATAATAATAATAATA 2416

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| |  |||||       || |||| |||
Db       2415 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2356

Qy        248 TAAAAAAATGCATTAGTTTG 267
                || |||| |  || ||||
Db       2355 ATAATAAATTCTGTATTTTG 2336


RESULT 9
US-11-097-143-16492
; Sequence 16492, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
```

```
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16492
;   LENGTH: 11143
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-16492

  Query Match             4.9%;  Score 51.6;  DB 10;  Length 11143;
  Best Local Similarity   52.9%;  Pred. No. 3.4;
  Matches  111;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy         68 ATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTG 127
              |||  |||||||||  | | |||| | | || ||||||  |  |||    | |  |  |
Db       4467 ATATAATATATAATATATAATATTATATTAATAATAATATTAATGATAATAATAATAATA 4526

Qy        128 AAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAA 187
              |  ||||   |    | |   || | |  ||      ||   | ||  ||||    || |
Db       4527 ATAATAATAATAATAATAATAATTATAATAATAACCATAATAATAATAATAATAATAATA 4586

Qy        188 ATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATT 247
              |||||||  | |||  | ||  ||||   ||| |  |||||      || |||| |||
Db       4587 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 4646

Qy        248 TAAAAAAATGCATTAGTTTGACGACAGGTA 277
                || ||    | || |   |  ||   ||
Db       4647 ATAATAATAATAATAATAACAGTACCAATA 4676


RESULT 10
US-10-437-963-98795
; Sequence 98795, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 98795
;   LENGTH: 441
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_9666C.1
US-10-437-963-98795

  Query Match             4.9%;  Score 51;  DB 7;  Length 441;
  Best Local Similarity   50.6%;  Pred. No. 1.2;
  Matches  123;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy         70 AGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAA 129
              ||||   ||||| |  || ||  |   || |||||| ||  | |    | |  |  | |
Db        128 AGCAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 187

Qy        130 GATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAAT 189
              ||||   |    | |  || | |  ||   || ||   | ||  ||||    || |||
Db        188 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 247

Qy        190 AATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTA 249
              |||||  | |||  | ||  ||||   ||| |  |||||       || |||| |||
Db        248 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 307

Qy        250 AAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTA 309
              || ||   | || |  |  |    |||  ||   |  | || | | ||  ||
Db        308 AATAATAATAATAATAATAATAATAATAATAATAATAATAATATTGTTATCCATTGTCGT 367

Qy        310 TTA 312
              ||
Db        368 CTA 370


RESULT 11
US-10-388-838-58
; Sequence 58, Application US/10388838
; Publication No. US20040180344A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001600
;  CURRENT APPLICATION NUMBER: US/10/388,838
;  CURRENT FILING DATE:  2003-03-14
;  NUMBER OF SEQ ID NOS: 114
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 58
;   LENGTH: 337344
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(337344)
;   OTHER INFORMATION: n = A,T,C or G
US-10-388-838-58

  Query Match             4.9%;  Score 50.8;  DB 8;  Length 337344;
  Best Local Similarity   46.2%;  Pred. No. 20;
  Matches  169;  Conservative    0;  Mismatches  197;  Indels    0;  Gaps    0;

Qy          5 ATGTTTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTAC 64
              || | | ||| | ||| | ||||   | | |  ||| |  |     | || |   | ||
Db     202599 ATATATATATAAATATATATATAATATATAAATATATATATATAAATATATATATAATAT 202658

Qy         65 TGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTT 124
                 |    ||||| ||  |    || | | |||| |   ||     |       || | |
Db     202659 ATAAATATATATACAAATATATATAATATATAAATATATATACAAATATATATATAATAT 202718

Qy        125 GTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAA 184
              | || ||| ||    | | | | || | ||| |      |||  |||  |||| |  |
Db     202719 ATAAATATATATACAAATATATATATAATATAAATATATATACAAATATATATATAAT 202778

Qy        185 AAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGA 244
              | |  |||| | | |    |     ||   |  || || |||  |     |||    | |
Db     202779 ATATAAATATATATACAAATATATATATAATATATAAATATATATACAAATATATATATA 202838

Qy        245 ATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATG 304
              || || ||| ||  ||   | |    | |     |  | |  ||   ||      |   ||
Db     202839 ATATATAAATATATATACATATATATATAATATATAAATATATATACAAATATATATATA 202898

Qy        305 ATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAAA 364
              || ||| |  | ||  |    |||  | |  |   | |  | |  | | |||    | |
Db     202899 ATATATAAATATATATACAAATATATATATAATATATAAATATATATACAAATATATATA 202958

Qy        365 TTGTAT 370
              |  |||
Db     202959 TAATAT 202964
```

```
RESULT 12
US-10-775-920-306
; Sequence 306, Application US/10775920
; Publication No. US20040175744A1
; GENERAL INFORMATION:
;  APPLICANT: Mergen Ltd
;  TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
;  TITLE OF INVENTION:  OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
;  TITLE OF INVENTION:  IN CERTAIN CANCERS
;  FILE REFERENCE: Mergen - 0010B
;  CURRENT APPLICATION NUMBER: US/10/775,920
;  CURRENT FILING DATE:  2004-02-10
;  PRIOR APPLICATION NUMBER: US 60/447,900
;  PRIOR FILING DATE: 2003-02-13
;  NUMBER OF SEQ ID NOS: 385
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 306
;   LENGTH: 4042
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-775-920-306

  Query Match             4.7%;  Score 48.8;  DB 8;  Length 4042;
  Best Local Similarity   57.1%;  Pred. No. 8.5;
  Matches   89;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         45 TGAGTTTTTAGAGCCACTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAA 104
              ||||  | |   |  | |||  |  |  ||    |  || ||||||| ||| || ||
Db       3882 TGAGGCTATGAATATAGTACCTGGAAAAATGCCAAGACATGATTATTTTTTTAAAAAGCG 3941

Qy        105 TGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTAT 164
              |  |||  | ||   ||||  | ||   |||  | |||| | |||||  | |||    |
Db       3942 TACTTTAAATGTTTGTGTTAATAAATTAAAACATGCACAAAAAGATGCAAAAAAAAAAAA 4001

Qy        165 TATCGAAAACTATAAAGCAAAAAATAATAAACAAAA 200
              |   ||||  | |||  |||||| || ||| ||||
Db       4002 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4037


RESULT 13
US-10-221-714A-66/c
; Sequence 66, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 66
;   LENGTH: 6179
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-66

  Query Match             4.7%;  Score 48.8;  DB 7;  Length 6179;
  Best Local Similarity   50.9%;  Pred. No. 10;
  Matches  116;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy         52 TTAGAGCCACTACTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTT 111
              | | | | |  |||  |    ||||| ||| || || | | |       ||| |   |
Db       3320 TAAAAACAAACACTACAACATATATAAAATAAAAATAAATATAATTTCCATATTACATAC 3261

Qy        112 GAGGTCATTGTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAA 171
                ||||  | | |      ||| |   | | | | || |   |   |  |||||  |
Db       3260 CTCATCATAATCTCTATTTTAAAAACAATATATATATATAATTAATTATAAATTATAAAT 3201

Qy        172 AACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTA 231
              || ||||||  | ||| ||| ||||||||  |     | |    ||| || | ||   |
Db       3200 AAATATAAATTATAAATTAAAAAACAAAAAATCACAAATCAAAAAAATTAACAAAAAAAA 3141

Qy        232 GGCTATGATAAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAA 279
                |   |||| ||  |||||||   || | | |  ||   ||||
Db       3140 AACACAAATAAAAAAAAAAAAAAAAAAAATAAATATTAAAACTACTAAA 3093


RESULT 14
US-10-027-632-119520
; Sequence 119520, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119520
;   LENGTH: 1127
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-119520

  Query Match             4.6%;  Score 48.4;  DB 5;  Length 1127;
  Best Local Similarity   51.1%;  Pred. No. 6.1;
  Matches  139;  Conservative    0;  Mismatches  131;  Indels    2;  Gaps    1;

Qy         16 AGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATA 75
              |  |  || ||||   | | |||||      | |  | | | |   | ||    |||  |||
Db        476 AATTATTATATAATATATAATAATTATTATAACATATAATAATAATTATAATATATAATA 535

Qy         76 TATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAA 135
              |||||| |  ||  |   | || ||| ||   ||  | | ||| ||   |  ||| |
Db        536 TATAATTATTATATTATATATAATAATTATTATATTATATATAATTATTATATATTATATA 595

Qy        136 TCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAA 195
              |   | |   | |  | ||||        ||   | || |||| |  | | |||||||||
Db        596 TAATAATAATTATAATATATAATAATATATAATAATAATTATATATTATATAATAATAAT 655

Qy        196 CAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAA 255
              | |     ||| | | |  || |  ||||||     ||  |||  |   | | | ||
Db        656 TATATATAATATATAATTAT--ATAATAATAATTATATATAATATATAATAATTATATAA 713
```

```
Qy        256 TGCATTAGTTTGACGACAGGTAAATATTGCAT 287
              | ||| | | | |  | |  |||  |||  ||
Db        714 TACATAATTATTATAATATATAATAATTATAT 745


RESULT 15
US-10-027-632-119520
; Sequence 119520, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119520
;   LENGTH: 1127
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-119520

  Query Match             4.6%;  Score 48.4;  DB 6;  Length 1127;
  Best Local Similarity   51.1%;  Pred. No. 6.1;
  Matches  139;  Conservative    0;  Mismatches  131;  Indels    2;  Gaps    1;

Qy         16 AGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATA 75
              | |  || ||||   | | |||||     | |  | | | |   | ||    |||  |||
Db        476 AATTATTATATAATATATAATAATTATTATAACATATAATAATAATTATAATATATAATA 535

Qy         76 TATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAA 135
              |||||| |  ||  |   |  || ||| ||  ||  |  | ||| ||    |  ||| |
Db        536 TATAATTATTATATTATATATAATAATTATATATTATATATAATTATTATATATTATATA 595

Qy        136 TCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAA 195
              |    | |   | |  | ||||       ||   | || |||| |  | | |||||||
Db        596 TAATAATAATTATAATATATAATAATATATAATAATAATTATATATTATATAATAATAAT 655

Qy        196 CAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAA 255
               | |     ||| | | | |  || |  ||||||     ||  |||  |   | | | ||
Db        656 TATATATAATATATAATTAT--ATAATAATAATTATATATAATATATAATAATTATATAA 713

Qy        256 TGCATTAGTTTGACGACAGGTAAATATTGCAT 287
              | ||| | | | |  | |  |||  |||  ||
Db        714 TACATAATTATTATAATATATAATAATTATAT 745


Search completed: May 12, 2006, 05:34:06
Job time : 673.458 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:52:36 ; Search time 531.917 Seconds
(without alignments)
8004.956 Million cell updates/sec

Title: US-09-423-093B-3_COPY_3977_5020
Perfect score: 1044
Sequence: 1 atgaatgtttttatcagtat.....ctccgcgtcttggtatttaa 1044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/2/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/2/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	52.2	5.0	1134	11	US-10-932-182A-166160	Sequence 166160,
2	52.2	5.0	1134	11	US-10-932-182A-166160	Sequence 166160,
3	48.4	4.6	966	12	US-10-301-480-553375	Sequence 553375,
4	48.4	4.6	966	12	US-10-301-480-1166784	Sequence 1166784,
5	48	4.6	752	11	US-10-301-480-27266	Sequence 27266, A
6	48	4.6	752	12	US-10-301-480-640675	Sequence 640675,
7	47	4.5	1086	11	US-10-932-182A-166166	Sequence 166166,
8	47	4.5	1086	11	US-10-932-182A-166166	Sequence 166166,
c 9	46.2	4.4	149382	10	US-10-995-561-13272	Sequence 13272, A
c 10	45.4	4.3	691	7	US-09-925-065A-758634	Sequence 758634,
c 11	45.4	4.3	691	7	US-09-925-065A-758635	Sequence 758635,
c 12	45.4	4.3	691	7	US-09-925-065A-758636	Sequence 758636,
13	45.2	4.3	1633	10	US-10-750-185-38087	Sequence 38087, A
14	45.2	4.3	1633	10	US-10-750-623-38087	Sequence 38087, A
15	45.2	4.3	319608	17	US-11-145-703-1	Sequence 1, Appli
16	45.2	4.3	1082144	17	US-11-117-187-211	Sequence 211, App
c 17	45	4.3	1045	7	US-09-925-065A-73230	Sequence 73230, A
c 18	45	4.3	1045	11	US-10-301-480-174469	Sequence 174469,
c 19	45	4.3	1045	12	US-10-301-480-787878	Sequence 787878,
20	44.8	4.3	544	12	US-10-301-480-551486	Sequence 551486,
21	44.8	4.3	544	12	US-10-301-480-1164895	Sequence 1164895,
22	44.8	4.3	1752	10	US-10-793-626-1357	Sequence 1357, Ap
23	44.8	4.3	3055	10	US-10-793-626-4075	Sequence 4075, Ap
24	44.4	4.3	2120	10	US-10-689-742-163	Sequence 163, App
25	44.4	4.3	184000	17	US-11-121-086-37	Sequence 37, Appl
c 26	43.6	4.2	488	7	US-09-925-065A-465917	Sequence 465917,
c 27	43.6	4.2	987	12	US-10-301-480-545289	Sequence 545289,
c 28	43.6	4.2	987	12	US-10-301-480-1158698	Sequence 1158698,
29	43.6	4.2	3143	17	US-11-165-819-9	Sequence 9, Appli
30	43.6	4.2	3143	17	US-11-165-819-29	Sequence 29, Appl
31	43.4	4.2	580	7	US-09-925-065A-761273	Sequence 761273,
32	43.4	4.2	580	7	US-09-925-065A-829838	Sequence 829838,
33	43.4	4.2	634	7	US-09-925-065A-487340	Sequence 487340,
34	43.4	4.2	634	7	US-09-925-065A-487341	Sequence 487341,
35	43.4	4.2	305312	10	US-10-995-561-13236	Sequence 13236, A
c 36	43.2	4.1	570	7	US-09-925-065A-433039	Sequence 433039,
c 37	43.2	4.1	630	7	US-09-925-065A-675278	Sequence 675278,
38	43.2	4.1	26772	10	US-10-995-561-13313	Sequence 13313, A
39	43.2	4.1	54946	10	US-10-995-561-13479	Sequence 13479, A
40	43.2	4.1	151169	17	US-11-121-086-38	Sequence 38, Appl
41	43	4.1	630	7	US-09-925-065A-607401	Sequence 607401,
42	42.8	4.1	1506	10	US-10-793-626-29	Sequence 29, Appl
43	42.8	4.1	1539	10	US-10-750-185-62749	Sequence 62749, A
44	42.8	4.1	1539	10	US-10-750-623-62749	Sequence 62749, A
45	42.8	4.1	3469	10	US-10-793-626-3409	Sequence 3409, Ap

ALIGNMENTS

RESULT 1
US-10-932-182A-166160
; Sequence 166160, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166160
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166160

Query Match 5.0%; Score 52.2; DB 11; Length 1134;
Best Local Similarity 46.2%; Pred. No. 4.3;
Matches 174; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

```
Qy        80 ATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTC 139
             || || || ||  |  ||  |  |  ||| | |   ||  |   | |   || | ||
Db       677 ATAAATATAATAATGGAATTATAACAAATTATCAACGTATATTAAATAATATTATACCTA 736

Qy       140 CACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAA 199
             |   | ||| | |  |  ||||| |   | |   ||||   ||| |||| ||| | |
Db       737 AATTAAATGATCATAATATTTCTATAAATTATATTAATAATATAAATAATATTAATAATA 796

Qy       200 ATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCA 259
             |  |||  |||| | | |  | || |    || |||| |     |||  ||  |
Db       797 TTAAATATAATAATATTATTACACCTTATGGAATTAATAATAATATAAATAAATTATATA 856
```

```
Qy        260 TTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTAGCAGATG 319
              || | | |  |  | |||  |  ||  ||  |  | |  ||||  || |
Db        857 ATAATATAAATATTGATAATATTCCTATAGATATTTTAATATTTAAATATTTAGTAGGTT 916

Qy        320 GAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAAATTGTATTGGCTACGC 379
              || |  |  | || ||  | | |  ||  || |  |  || |           ||
Db        917 GATCTATTAAATTAAAAGGTAGATTAAGTAATAATATTGGTAGAACTAGTACTCTTATTT 976

Qy        380 GAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGTGATACTGTTCGTCATTTAA 439
              |  |||||  |||||  | ||  ||  | | || ||| |  |  | |||  |
Db        977 TATTAAATGGTACTTTTAATAATAAAAAATATTTATGAAGTAATATTAATAATAATTATA 1036

Qy        440 CAGACGATACTTTATTT 456
              |  ||  | | | |
Db       1037 AATTAAATTATATCTCT 1053
```

RESULT 2
US-10-932-182A-166160
; Sequence 166160, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166160
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166160

Query Match 5.0%; Score 52.2; DB 11; Length 1134;
Best Local Similarity 46.2%; Pred. No. 4.3;
Matches 174; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

```
Qy         80 ATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTC 139
              || || ||  ||  |  ||  |  |  |||  | |  ||  |  | |  || | ||
Db        677 ATAAATATAATAATGGAATTATAACAAATTATCAACGTATATTAAATAATATTATACCTA 736

Qy        140 CACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAA 199
              |    | ||| | |  |  |||||  |   | |   ||||   ||| |||| |||  | |
Db        737 AATTAAATGATCATAATATTTCTATAAATTATATTAATAATATAAATAATATTAATAATA 796

Qy        200 ATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCA 259
              |   |||  |||| | | |  | | ||  |   || ||||  |     |||  ||  |
Db        797 TTAAATATAATAATATTATTACACCTTATGGAATTAATAATAATATAAATAAATTATATA 856

Qy        260 TTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGATGATCTATTAGCAGATG 319
              || | | |  |  | |||  |  ||  ||  |  | |  ||||  || |
Db        857 ATAATATAAATATTGATAATATTCCTATAGATATTTTAATATTTAAATATTTAGTAGGTT 916

Qy        320 GAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAAATTGTATTGGCTACGC 379
              || |  |  | || ||  | | |  ||  || |  |  || |           ||
Db        917 GATCTATTAAATTAAAAGGTAGATTAAGTAATAATATTGGTAGAACTAGTACTCTTATTT 976

Qy        380 GAGCGTATGGTTGGTTTAAGGAAAATCCGAATGAGTTATGTGATACTGTTCGTCATTTAA 439
              |  |||||  |||||  | ||  ||  | | || ||| |  |  | |||  |
Db        977 TATTAAATGGTACTTTTAATAATAAAAAATATTTATGAAGTAATATTAATAATAATTATA 1036

Qy        440 CAGACGATACTTTATTT 456
              |  ||  | | | |
Db       1037 AATTAAATTATATCTCT 1053
```

RESULT 3
US-10-301-480-553375
; Sequence 553375, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 553375
; LENGTH: 966
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-553375

Query Match 4.6%; Score 48.4; DB 12; Length 966;
Best Local Similarity 51.1%; Pred. No. 20;
Matches 139; Conservative 0; Mismatches 131; Indels 2; Gaps 1;

```
Qy         16 AGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATA 75
              |  || |||| | | |||||  |  |  | | | |  | ||  |||  |||
Db        454 AATTATTATATAATATATAATAATTATTATAACATATAATAATAATTATAATATATAATA 513

Qy         76 TATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAA 135
              |||||| |  ||  |   |  || ||| ||   ||  |  | ||| ||     |  ||| |
Db        514 TATAATTATTATATTATATATAATAATTATATATTATATATAATTATTATATATTATATA 573

Qy        136 TCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAA 195
              |    | |  | |  | ||||      ||   | || |||| | | | ||||||||
Db        574 TAATAATAATTATAATATATAATAATATATAATAATAATTATATATTATATAATAATAAT 633

Qy        196 CAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAA 255
              | |      ||| | | | |  || |  ||||||     ||  |||  |   | | | ||
Db        634 TATATATAATATATAATTAT--ATAATAATAATTATATATAATATATAATAATTATATAA 691

Qy        256 TGCATTAGTTTGACGACAGGTAAATATTGCAT 287
              | ||| | | | |  | |  |||  |||  ||
Db        692 TACATAATTATTATAATATATAATAATTATAT 723
```

RESULT 4
US-10-301-480-1166784
; Sequence 1166784, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1166784
; LENGTH: 966
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-1166784

Query Match 4.6%; Score 48.4; DB 12; Length 966;
Best Local Similarity 51.1%; Pred. No. 20;

```
  Matches  139;  Conservative    0;  Mismatches  131;  Indels    2;  Gaps    1;

Qy          16 AGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATA 75
               | |  || ||||   | | |||||     | |  | | | |   | ||   |||  |||
Db         454 AATTATTATATAATATATAATAATTATTATAACATATAATAATAATTATAATATATAATA 513

Qy          76 TATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAA 135
               ||||||  |  ||  |   |  || ||| ||   ||  |  | ||| ||    |  ||| |
Db         514 TATAATTATTATATTATATATAATAATTATATATTATATATAATTATTATATATTATATA 573

Qy         136 TCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAA 195
               |    | |   | |  | ||||       ||   | || |||| |  | | ||||||||
Db         574 TAATAATAATTATAATATATAATAATATATAATAATAATTATATATTATATAATAATAAT 633

Qy         196 CAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAA 255
                | |     ||| | | | |  || |  ||||||     ||  |||  |   | | | ||
Db         634 TATATATAATATATAATTAT--ATAATAATAATTATATATAATATATAATAATTATATAA 691

Qy         256 TGCATTAGTTTGACGACAGGTAAATATTGCAT 287
               | ||| | | | |  | |  |||  |||  ||
Db         692 TACATAATTATTATAATATATAATAATTATAT 723


RESULT 5
US-10-301-480-27266
; Sequence 27266, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27266
;   LENGTH: 752
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-27266

  Query Match             4.6%;  Score 48;  DB 11;  Length 752;
  Best Local Similarity   50.7%;  Pred. No. 22;
  Matches  138;  Conservative    1;  Mismatches  131;  Indels    2;  Gaps    1;

Qy          16 AGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATA 75
               | |  || ||||   | | |||||     | |  | | | |   | ||   |||  |||
Db         100 AATTATTATATAATATATAATAATTATTATAACATATAATAATAATTATAATATATAATA 159

Qy          76 TATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAA 135
               ||||||  |  ||  |   |  || ||| ||   ||  |  | ||| ||    |  ||| |
Db         160 TATAATTATTATATTATATATAATAATTATATATTATATATAATTATTATATATTATATA 219

Qy         136 TCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAA 195
               |    | |   | :  | ||||       ||   | || |||| |  | | ||||||||
Db         220 TAATAATAATTATWATATATAATAATATATAATAATAATTATATATTATATAATAATAAT 279

Qy         196 CAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAA 255
                | |     ||| | | | |  || |  ||||||     ||  |||  |   | | | ||
Db         280 TATATATAATATATAATTAT--ATAATAATAATTATATATAATATATAATAATTATATAA 337

Qy         256 TGCATTAGTTTGACGACAGGTAAATATTGCAT 287
               | ||| | | | |  | |  |||  |||  ||
Db         338 TACATAATTATTATAATATATAATAATTATAT 369


RESULT 6
US-10-301-480-640675
; Sequence 640675, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 640675
;   LENGTH: 752
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-640675

  Query Match             4.6%;  Score 48;  DB 12;  Length 752;
  Best Local Similarity   50.7%;  Pred. No. 22;
  Matches  138;  Conservative    1;  Mismatches  131;  Indels    2;  Gaps    1;

Qy          16 AGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGATAGCATA 75
               | |  || ||||   | | |||||     | |  | | | |   | ||   |||  |||
Db         100 AATTATTATATAATATATAATAATTATTATAACATATAATAATAATTATAATATATAATA 159

Qy          76 TATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGAAGATAAA 135
               ||||||  |  ||  |   |  || ||| ||   ||  |  | ||| ||    |  ||| |
Db         160 TATAATTATTATATTATATATAATAATTATATATTATATATAATTATTATATATTATATA 219

Qy         136 TCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAA 195
               |    | |   | :  | ||||       ||   | || |||| |  | | ||||||||
Db         220 TAATAATAATTATWATATATAATAATATATAATAATAATTATATATTATATAATAATAAT 279

Qy         196 CAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAA 255
                | |     ||| | | | |  || |  ||||||     ||  |||  |   | | | ||
Db         280 TATATATAATATATAATTAT--ATAATAATAATTATATATAATATATAATAATTATATAA 337

Qy         256 TGCATTAGTTTGACGACAGGTAAATATTGCAT 287
               | ||| | | | |  | |  |||  |||  ||
Db         338 TACATAATTATTATAATATATAATAATTATAT 369


RESULT 7
US-10-932-182A-166166
; Sequence 166166, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAO, YOSHIHIRO
;  APPLICANT:  NAKAMURA, NORIHISA
;  APPLICANT:  KODAMA, YUKIKO
;  APPLICANT:  FUJIMURA, TOMOKO
;  APPLICANT:  ASHIKARI, TOSHIHIKO
;  TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
;  FILE REFERENCE: 030685-043
;  CURRENT APPLICATION NUMBER: US/10/932,182A
;  CURRENT FILING DATE:  2004-09-02
;  NUMBER OF SEQ ID NOS: 197023
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166166
;   LENGTH: 1086
;   TYPE: DNA
;   ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166166

  Query Match             4.5%;  Score 47;  DB 11;  Length 1086;
  Best Local Similarity   46.1%;  Pred. No. 35;
  Matches  158;  Conservative    0;  Mismatches  185;  Indels    0;  Gaps    0;

Qy          64 CTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTT 123
```

```
Db        610 CTATATTTTTTATATTATGTTATTATTCCTAAATTTAATAATATAAATTGATATACTATT 669

Qy        124 TGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCA 183

Db        670 AAATATAATAGTTTTATAAATTGACAGTCTATTATTAATATTTATATTAAAGGTTTACAT 729

Qy        184 AAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAG 243

Db        730 AAAATTAATAATAATAATGTTATTCAAGAATATATTGAACATCTTATAAATTTAAATAAA 789

Qy        244 AATTTAAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGAT 303

Db        790 AAAGGTGCTAATAAAATTAATTATAATATTGAATTATATAATAAAATTATTAATAATAAT 849

Qy        304 GATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAA 363

Db        850 AAACCATTATATAATATGAATTTATCATATAAACTTAATAGTTTAAATTATAAAACTAAT 909

Qy        364 ATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATC 406

Db        910 AAGTTTATTGGTGTAGGAGTATATGTATATGATTTAAATAATC 952
```

RESULT 8
US-10-932-182A-166166
; Sequence 166166, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166166
; LENGTH: 1086
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166166

Query Match 4.5%; Score 47; DB 11; Length 1086;
Best Local Similarity 46.1%; Pred. No. 35;
Matches 158; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

```
Qy         64 CTGGATAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTT 123

Db        610 CTATATTTTTTATATTATGTTATTATTCCTAAATTTAATAATATAAATTGATATACTATT 669

Qy        124 TGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCA 183

Db        670 AAATATAATAGTTTTATAAATTGACAGTCTATTATTAATATTTATATTAAAGGTTTACAT 729

Qy        184 AAAAATAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAG 243

Db        730 AAAATTAATAATAATAATGTTATTCAAGAATATATTGAACATCTTATAAATTTAAATAAA 789

Qy        244 AATTTAAAAAAAATGCATTAGTTTGACGACAGGTAAATATTGCATGATCATGGGCAACGAT 303

Db        790 AAAGGTGCTAATAAAATTAATTATAATATTGAATTATATAATAAAATTATTAATAATAAT 849

Qy        304 GATCTATTAGCAGATGGAGCGTTATCAAAAATAGTGAAAGTTTTGAAGGCTAATCCTGAA 363

Db        850 AAACCATTATATAATATGAATTTATCATATAAACTTAATAGTTTAAATTATAAAACTAAT 909

Qy        364 ATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGGAAAATC 406

Db        910 AAGTTTATTGGTGTAGGAGTATATGTATATGATTTAAATAATC 952
```

RESULT 9
US-10-995-561-13272/c
; Sequence 13272, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13272
; LENGTH: 149382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(149382)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-995-561-13272

Query Match 4.4%; Score 46.2; DB 10; Length 149382;
Best Local Similarity 50.3%; Pred. No. 98;
Matches 99; Conservative 5; Mismatches 93; Indels 0; Gaps 0;

```
Qy        160 TCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAAT 219

Db      46596 TCCAAAAACTAAAGATTAAAATCAACAAATTACATACAATATGCCCATAATAAGAACAAT 46537

Qy        220 GAAGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAA 279

Db      46536 CATAATAATTTACAYAATKATGTCCCCTTTAAACACCAGGTAAYATCCAAAAAACTTGAG 46477

Qy        280 TATTGCATGATCATGGGCAACGATGATCTATTAGCAGATGGAGCGTTATCAAAAATAGTG 339

Db      46476 AACCCCATACCCAYRGGCCAAGGYTTCTCCTAAATAGATGTGGGTATCTCTAATTTAATA 46417

Qy        340 AAAGTTTTGAAGGCTAA 356

Db      46416 TAAATGAGGAAAGAAAA 46400
```

RESULT 10
US-09-925-065A-758634/c
; Sequence 758634, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 758634
; LENGTH: 691
; TYPE: DNA

```
;   ORGANISM: Homo sapiens
US-09-925-065A-758634

  Query Match             4.3%;  Score 45.4;  DB 7;  Length 691;
  Best Local Similarity   65.0%;  Pred. No. 64;
  Matches   67;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy        154 ATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAAT 213
              ||||  |  || ||  | ||| ||||   || ||||||||||| ||||| |  || |  ||
Db        159 ATAATTTTAATAATTAATAACAATAATATAATAAATAATAAATAAAATATAAATATATAT 100

Qy        214 TTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAAT 256
              || | |  | |||||  |    || |||| ||||| || || |
Db         99 TTTATTATACATAATAAATAAAATAATAATAATTTTAATAACT 57


RESULT 11
US-09-925-065A-758635/c
; Sequence 758635, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 758635
;   LENGTH: 691
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-758635

  Query Match             4.3%;  Score 45.4;  DB 7;  Length 691;
  Best Local Similarity   65.0%;  Pred. No. 64;
  Matches   67;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy        154 ATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAAT 213
              ||||  |  || ||  | ||| ||||   || ||||||||||| ||||| |  || |  ||
Db        159 ATAATTTTAATAATTAATAACAATAATATAATAAATAATAAATAAAATATAAATATATAT 100

Qy        214 TTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAAT 256
              || | |  | |||||  |    || |||| ||||| || || |
Db         99 TTTATTATACATAATAAATAAAATAATAATAATTTTAATAACT 57


RESULT 12
US-09-925-065A-758636/c
; Sequence 758636, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 758636
;   LENGTH: 691
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-758636

  Query Match             4.3%;  Score 45.4;  DB 7;  Length 691;
  Best Local Similarity   65.0%;  Pred. No. 64;
  Matches   67;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy        154 ATAAACTCTATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAAT 213
              ||||  |  || ||  | ||| ||||   || ||||||||||| ||||| |  || |  ||
Db        159 ATAATTTTAATAATTAATAACAATAATATAATAAATAATAAATAAAATATAAATATATAT 100

Qy        214 TTCAATGAAGATAATTTAGGCTATGATAAGAATTTAAAAAAAT 256
              || | |  | |||||  |    || |||| ||||| || || |
Db         99 TTTATTATACATAATAAATAAAATAATAATAATTTTAATAACT 57


RESULT 13
US-10-750-185-38087
; Sequence 38087, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 38087
;   LENGTH: 1633
;   TYPE: DNA
;   ORGANISM: Bovine  19866880592950
US-10-750-185-38087

  Query Match             4.3%;  Score 45.2;  DB 10;  Length 1633;
  Best Local Similarity   47.2%;  Pred. No. 78;
  Matches  169;  Conservative    0;  Mismatches  188;  Indels    1;  Gaps    1;

Qy        102 TAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTC 161
              |||  | | || ||  ||       |  |  | |   || ||| | | || | ||| |
Db       1159 TAAATACTCTGTGGCAATAAGGCCTTTAACTCATATTCTACAAATACATAAAATATAACT 1218

Qy        162 TATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGA 221
              ||  || |||     ||    | | | | |  ||| | |  |  |   |||||| | |||
Db       1219 TAAAATTGAACTAATTATGCTACATATTGAATAACTATAAATAAAATATAATTTGAGTGA 1278

Qy        222 AGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATA 281
              |  |   | | |  | | ||| |||    ||||   || |      ||||      |
Db       1279 ACTTCCATAATGAATTTAGAAGCATTCTTTAAAACTGATGATAAATGTGACACTGCTGTT 1338

Qy        282 TTGCATGATCATGGGCAACGATGATCTATTAGC-AGATGGAGCGTTATCAAAAATAGTGA 340
              |      |  || |  || || | || | ||   ||||    || || | |||  |||
Db       1339 TAAATGAAGTCTGAGACACTATCAGCTGTAAGAGTGATGCTTTGTCATAACAAACTGTGT 1398
```

```
Qy        341 AAGTTTTGAAGGCTAATCCTGAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGG 400
              ||   |  ||  ||||   | |  | | |   ||      || |   |||||||  |
Db       1399 AAAAATATAAATCTAAGGTTTAGGTGGAAGAAGCACTTTGACCAGCTGGTTGGGATTCAA 1458

Qy        401 AAAATCCGAATGAGTTATGTGATACTGTTCGTCATTTAACAGACGATACTTTATTTCA 458
                  | | ||  | | | |  | |||||||| | |    |  |||| ||
Db       1459 GTCTAGGTCTTTTGCTACATTAAATTTATTGTCATTTAAAATTCTCATCCATATTGCA 1516


RESULT 14
US-10-750-623-38087
; Sequence 38087, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-1
;  CURRENT APPLICATION NUMBER: US/10/750,623
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 38087
;   LENGTH: 1633
;   TYPE: DNA
;   ORGANISM: Bovine  19866880592950
US-10-750-623-38087

  Query Match             4.3%;  Score 45.2;  DB 10;  Length 1633;
  Best Local Similarity   47.2%;  Pred. No. 78;
  Matches  169;  Conservative    0;  Mismatches  188;  Indels    1;  Gaps    1;

Qy        102 TAATGATTTTGAGGTCATTGTTTGTGAAGATAAATCTCCACAGAGAGATGAGATAAACTC 161
              |||  | | ||  ||  ||       | |  | |   || ||| | | || | ||| |
Db       1159 TAAATACTCTGTGGCAATAAGGCCTTTAACTCATATTCTACAAATACATAAAATATAACT 1218

Qy        162 TATTATCGAAAACTATAAAGCAAAAAATAATAAACAAAATCTTTATGTTAATTTCAATGA 221
              ||  || |||     ||    | | | | |  ||| | |  |  |   |||||| | |||
Db       1219 TAAAATTGAACTAATTATGCTACATATTGAATAACTATAAATAAAATATAATTTGAGTGA 1278

Qy        222 AGATAATTTAGGCTATGATAAGAATTTAAAAAAATGCATTAGTTTGACGACAGGTAAATA 281
              |  |   | | |   | | ||| |||    ||||   || |       ||||      |
Db       1279 ACTTCCATAATGAATTTAGAAGCATTCTTTAAAACTGATGATAAATGTGACACTGCTGTT 1338

Qy        282 TTGCATGATCATGGGCAACGATGATCTATTAGC-AGATGGAGCGTTATCAAAAATAGTGA 340
              |       |   || |  || || | || | ||   ||||    || || | |||  |||
Db       1339 TAAATGAAGTCTGAGACACTATCAGCTGTAAGAGTGATGCTTTGTCATAACAAACTGTGT 1398

Qy        341 AAGTTTTGAAGGCTAATCCTGAAATTGTATTGGCTACGCGAGCGTATGGTTGGTTTAAGG 400
              ||   |  ||  ||||   | |  | | |   ||      || |   |||||||  |
Db       1399 AAAAATATAAATCTAAGGTTTAGGTGGAAGAAGCACTTTGACCAGCTGGTTGGGATTCAA 1458

Qy        401 AAAATCCGAATGAGTTATGTGATACTGTTCGTCATTTAACAGACGATACTTTATTTCA 458
                  | | ||  | | | |  | |||||||| | |    |  |||| ||
Db       1459 GTCTAGGTCTTTTGCTACATTAAATTTATTGTCATTTAAAATTCTCATCCATATTGCA 1516


RESULT 15
US-11-145-703-1
; Sequence 1, Application US/11145703
; Publication No. US20050260667A1
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE: 53.US16.DIV
;  CURRENT APPLICATION NUMBER: US/11/145,703
;  CURRENT FILING DATE:  2000-03-30
;  PRIOR APPLICATION NUMBER: US/10/147,603
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 09/539,333
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: US 60/126,903
;  PRIOR FILING DATE: 1999-03-30
;  PRIOR APPLICATION NUMBER: US 60/131,971
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/132,065
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/143,928
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 60/145,915
;  PRIOR FILING DATE: 1999-07-27
;  PRIOR APPLICATION NUMBER: US 60/146,453
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/146,452
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/162,288
;  PRIOR FILING DATE: 1999-10-28
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 234
;  SOFTWARE: Patent.pm
; SEQ ID NO 1
;   LENGTH: 319608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 31..1107
;   OTHER INFORMATION: 5'regulatory region g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 1108..1289
;   OTHER INFORMATION: exon A g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 14877..14920
;   OTHER INFORMATION: exon B g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 18778..18862
;   OTHER INFORMATION: exon Bbis g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 25593..25740
;   OTHER INFORMATION: exon C g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 29388..29502
;   OTHER INFORMATION: exon D g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 29967..30282
;   OTHER INFORMATION: exon E g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 64666..64812
;   OTHER INFORMATION: exon F g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 65505..65853
;   OTHER INFORMATION: exon G g35018 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 65854..67854
;   OTHER INFORMATION: 3'regulatory region g35018 gene
```

; FEATURE:
; NAME/KEY: exon
; LOCATION: 94124..94964
; OTHER INFORMATION: exon g35017
; FEATURE:
; NAME/KEY: exon
; LOCATION: 201188..201234
; OTHER INFORMATION: exon S g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: exon T g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216836..216915
; OTHER INFORMATION: exon V g35030 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 213818..215818
; OTHER INFORMATION: 3'regulatory region g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215941
; OTHER INFORMATION: exon R complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215975
; OTHER INFORMATION: exon Rbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..216952
; OTHER INFORMATION: exon Qbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..217061
; OTHER INFORMATION: exon Q complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 217027..217061
; OTHER INFORMATION: exon Q1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 229647..229742
; OTHER INFORMATION: exon X complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 230408..230721
; OTHER INFORMATION: exon P complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231272..231412
; OTHER INFORMATION: exon Obis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231787..231880
; OTHER INFORMATION: exon O2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231870..231879
; OTHER INFORMATION: exon O1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 234174..234321
; OTHER INFORMATION: exon O complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 237406..237428
; OTHER INFORMATION: exon Nbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 239719..239807
; OTHER INFORMATION: exon N2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 239719..239853
; OTHER INFORMATION: exon N complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240569
; OTHER INFORMATION: exon M1117 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240596
; OTHER INFORMATION: exon M1090 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240617
; OTHER INFORMATION: exon M1069 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240644
; OTHER INFORMATION: exon MS2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240824
; OTHER INFORMATION: exon M862 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240994
; OTHER INFORMATION: exon M692 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..241685
; OTHER INFORMATION: exon M1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240800..240993
; OTHER INFORMATION: exon MS1 complement g34872 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 241686..243685
; OTHER INFORMATION: 5'regulatory region g34872 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 290652..292652
; OTHER INFORMATION: 3'regulatory region g34665 gene

Query Match 4.3%; Score 45.2; DB 17; Length 319608;
Best Local Similarity 48.4%; Pred. No. 1.6e+02;
Matches 125; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

```
Qy          9 TTTTATCAGTATTTGTATACCGTCTTATAATAGAGCTGAGTTTTTAGAGCCACTACTGGA 68
              || | |   |||    |||   | ||||          | | | || |     ||    |
Db     187708 TTATTTATATATAAAAATAATATATTATTTATATATATAATATATATATTATATATATAA 187767

Qy         69 TAGCATATATAATCAAGATTATTGTTTAAAGAATAATGATTTTGAGGTCATTGTTTGTGA 128
              ||  |||||||||  | | |||    || | ||||      |       ||| ||| |
Db     187768 TAATATATATAATATATAATATATAATATATAATATAATAATAATAAATATTATTTATAT 187827

Qy        129 AGATAAATCTCCACAGAGAGATGAGATAAACTCTATTATCGAAAACTATAAAGCAAAAAA 188
              | |||||| |  | | | | ||   |||  || ||| |    |||     ||  |||  |
Db     187828 ATATAAATTTATATATATAAATTTAATATGCTATATAAATATAAATATATAAATAAAGTA 187887

Qy        189 TAATAAACAAAATCTTTATGTTAATTTCAATGAAGATAATTTAGGCTATGATAAGAATTT 248
              || |  ||||   | | ||   |||||  | |  |     ||||  |     | ||||
Db     187888 TACTCCACAATTGCCTAATACAGATTTCTTTTACAAATCCATAGGACACCTATAAAATTG 187947

Qy        249 AAAAAAATGCATTAGTTT 266
              | | | ||  | |   ||
Db     187948 ACAGATATAAAATGTGTT 187965
```

Search completed: May 12, 2006, 06:54:58
Job time : 533.917 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 08:31:04 ; Search time 3638.94 Seconds
(without alignments)
15792.702 Million cell updates/sec

Title: US-09-423-093B-3_COPY_6313_7323
Perfect score: 1011
Sequence: 1 atgaataataaaaaggtttt.....catgtgatggaaataaatga 1011

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1011	100.0	12441	1	SERFBC2	X61917 S.enterica
	2	1011	100.0	12441	6	BD085556	BD085556 Nucleic a
	3	1011	100.0	12446	6	AX029263	AX029263 Sequence
	4	270.6	26.8	1029	6	AR384258	AR384258 Sequence
	5	114.4	11.3	110000	1	BA000016_25	Continuation (26 o
	6	113.6	11.2	110000	1	AJ749949_08	Continuation (9 of
c	7	112.6	11.1	12927	1	AE007804	AE007804 Clostridi
	8	110.8	11.0	110000	1	BA000019_53	Continuation (54 o
	9	106	10.5	14693	1	AE000693	AE000693 Aquifex a
c	10	102.4	10.1	110000	1	CP000099_02	Continuation (3 of
	11	100.2	9.9	11768	1	AE013289	AE013289 Methanosa
	12	100	9.9	10745	1	AE007788	AE007788 Clostridi
	13	98.6	9.8	15890	1	AB046360	AB046360 Actinobac
c	14	97.4	9.6	11501	1	AE009578	AE009578 Brucella
c	15	97.4	9.6	13835	1	AF047478	AF047478 Brucella
	16	97.4	9.6	110000	1	AE014291_05	Continuation (6 of
	17	97.4	9.6	110000	1	AE017223_05	Continuation (6 of
c	18	93.4	9.2	110000	1	CP000095_11	Continuation (12 o
	19	88.8	8.8	110000	1	CR543861_00	CR543861 Acinetoba
c	20	87.8	8.7	110000	1	BA000022_16	Continuation (17 o
c	21	85.6	8.5	300029	1	AE015936	AE015936 Clostridi
	22	84.4	8.3	65865	1	AY211269	AY211269 Campyloba
c	23	84.2	8.3	110000	1	BA000022_26	Continuation (27 o
	24	83	8.2	10029	1	AE011087	AE011087 Methanosa
c	25	82.2	8.1	10029	1	AE013290	AE013290 Methanosa
	26	80.8	8.0	13403	1	AY693427	AY693427 Shigella
	27	80.8	8.0	14702	1	AF402314	AF402314 Shigella
c	28	80.8	8.0	110000	1	BA000043_33	Continuation (34 o
	29	80	7.9	110000	1	BA000022_24	Continuation (25 o
	30	78.8	7.8	11469	1	AE011627	AE011627 Xanthomon
	31	78.8	7.8	13144	1	AF189151	AF189151 Klebsiell
	32	78.8	7.8	13316	1	ECORFBM	D43637 Escherichia
	33	75.8	7.5	213732	1	AE001862	AE001862 Deinococc
	34	75	7.4	15295	1	AB010150	AB010150 Escherich
	35	74	7.3	1161	6	AR384424	AR384424 Sequence
	36	73.8	7.3	110000	1	BA000045_41	Continuation (42 o
c	37	73.2	7.2	61985	3	AY744396	AY744396 Unculture
c	38	73.2	7.2	110000	1	AP008231_22	Continuation (23 o
c	39	72.4	7.2	110000	1	AP006841_44	Continuation (45 o
	40	66.4	6.6	110000	1	BA000037_03	Continuation (4 of
	41	65.6	6.5	21277	1	AF343089	AF343089 Aeromonas
	42	65.6	6.5	21279	1	AF148126	AF148126 Aeromonas
	43	65.2	6.4	110000	1	AE017261_12	Continuation (13 o
	44	64.8	6.4	16033	1	AE000749	AE000749 Aquifex a
	45	64.8	6.4	110000	1	BA000022_03	Continuation (4 of

ALIGNMENTS

RESULT 1
SERFBC2
LOCUS SERFBC2 12441 bp DNA linear BCT 09-SEP-2004
DEFINITION S.enterica rfbJ gene cluster.
ACCESSION X61917
VERSION X61917.1 GI:47004
KEYWORDS abe gene; abequose synthase; galactosyltransferase; manB gene; manC gene; mannose-1-phosphate guanyltransferase; phosphomannomutase; wbaL gene; wbaP gene; wbaQ gene; wbaR gene; wbaW gene; wbaZ gene; wzx gene; wzy gene.
SOURCE Salmonella enterica
ORGANISM Salmonella enterica
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 12441)
AUTHORS Brown,P.K., Romana,L.K. and Reeves,P.R.
TITLE Molecular analysis of the rfb gene cluster of Salmonella serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B
JOURNAL Mol. Microbiol. 6 (10), 1385-1394 (1992)
PUBMED 1379320
REFERENCE 2
AUTHORS Reeves,P.
TITLE Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale
JOURNAL Trends Genet. 9 (1), 17-22 (1993)
PUBMED 8434412
REFERENCE 3
AUTHORS Liu,D., Haase,A.M., Lindqvist,L., Lindberg,A.A. and Reeves,P.R.
TITLE Glycosyl transferases of O-antigen biosynthesis in Salmonella enterica: identification and characterization of transferase genes of groups B, C2, and E1
JOURNAL J. Bacteriol. 175 (11), 3408-3413 (1993)
PUBMED 7684736
REFERENCE 4 (bases 1 to 12441)
AUTHORS Brown,P.K.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1991) P.K. Brown, University of Sydney, Dept of Microbiology, Maze Crescent, Sydney N S W 2006, AUSTRALIA
FEATURES Location/Qualifiers
source 1. .12441

/organism="Salmonella enterica"
/mol_type="genomic DNA"
/strain="group C2, serovar muenchen M67"
/db_xref="taxon:28901"
gene 138. .1019
/gene="abe"
CDS 138. .1019
/gene="abe"
/codon_start=1
/transl_table=11
/product="abequose synthase"
/protein_id="CAA43918.1"
/db_xref="GI:47005"
/db_xref="GOA:Q00329"
/db_xref="UniProt/Swiss-Prot:Q00329"
/translation="MLDVNKKILMTGATSFVGTHLLHSLIKEGYSIIALKRPITEPTI
INTLIEWLNIQDIEKICQSSMNIHAIVHIATDYGRNRTPISEQYKCNVLLPTRLLELM
PALKTKFFISTDSFFGKYEKHYGYMRSYMASKRHFVELSKIYVEEHPDVCFINLRLEH
VYGERDKAGKIIPYVIKKMKNNEDIDCTIARQKRDFIYIDDVVSAYLKILKEGFNAGH
YDVEVGTGKSIELKEVFEIIKKETHSSSKINYGAVAMRDDEIMESHANTSFLTRLGWS
AEFSIEKGVKKMLSMKE"
CDS 1019. .2359
/note="wzx
Referred to as orf12.6 in reference [2]"
/citation=[2]
/codon_start=1
/transl_table=11
/product="hypothetical protein with 12 predicted
transmembrane segments which possibly functions in export
of O-antigen"
/protein_id="CAA43919.1"
/db_xref="GI:47006"
/db_xref="UniProt/TrEMBL:Q00339"
/translation="MNRIIRMLGVDKAIRYVIFGKIISVLTGLLLIMLISHHLSKDAQ
GYYYTFNSVVALQIIFELGLSTVIIQFASHEMSALKYDYSERDIIGESKNKQRYLSLF
RLAIKWYAVIALLIILIVGPIGYVFFTQKEGLGVPWQGAWLLLTIVTAFNIFLVSVLS
VAEGSGLITDVNKMRMYQSLLAGILAVSLLISGFGLYATSAIAISGTIIFSIFSYKYF
KKIFLQSLKHKNKYTEGGISWVNEIFPMQWRIALSWMSGYFIYFVMTPIAFKYFGAIY
AGQLGMSLTLCNMVMATGLAWISTKYPKWGVMVSNKQLAELSKSFKSAVMQSSFFVLT
GLTGVYISLWLLKLSGSNIGERFLGLQDFFFLSLAIIGNHIVACFATYIRAHKTEKMT
LASCIMALLTITTMLFVAYLEYSRFYMLMYAALTWLYFVPQTYIIFKRFKSSYE"
CDS 2352. .3314
/note="unnamed protein product; wbaR"
/codon_start=1
/transl_table=11
/protein_id="CAA43909.1"
/db_xref="GI:47007"
/db_xref="UniProt/TrEMBL:Q00337"
/translation="MSKKPLLTIAIPTYNRSSCLARLLDSIIQQENYCHDELEVIVCD
NASTDETARIAKSGLDKIRNSTYHLNEENLGMDGNFQKCFELSNGKYLWMIGDDDLIV
KNGISKVFSILKSRPALDMVYVNSAAKTELNYNADVRTSFYTNDVDFISDVKVMFTFI
SGMICKKTDAIVKAVGIFSPQTTGKYLMHLTWQLPLLKQGGEFAVIHNNIIEAEPDNS
GGYHLYKVFSNNLATIFDVFYPREHRVSKRVRASACLFLLNFIGDEDKTKNFATNNYL
RDCDSAFIDLIIYKYGLRFFYLYPKTVPLFRKIKYIIKTVLMRK"
CDS 3361. .3975
/note="unnamed protein product; wbaL
orf14.9 possibly encodes an acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="CAA43910.1"
/db_xref="GI:47008"
/db_xref="UniProt/TrEMBL:Q00480"
/translation="MFVYSLRLKLNLIISLLSKVRRKSKAKFLVLLSGYDFKMVGKNF
KLNVKPYSAKNNTSSKWGSMRVGDNCWIEAVYNYGDEKFEPYLYIGDRICLSDNVHIS
CVSCLILENDILIGSKVYIGDHSHGSYKVCSPKIEPPANKPLGDIAPIKIGNCCWIGD
NAVILAGSEICDGCVIAANSVVKDLKVDKPCLIGGVPAKVIKVF"
CDS 3977. .5020
/note="wbaQ
Referred to as orf15.6 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="rhamnosyl transferase"
/protein_id="CAA43911.1"
/db_xref="GI:47009"
/db_xref="GOA:Q00482"
/db_xref="UniProt/TrEMBL:Q00482"
/translation="MNVFISICIPSYNRAEFLEPLLDSIYNQDYCLKNNDFEVIVCED
KSPQRDEINSIIENYKAKNNKQNLYVNFNEDNLGYDKNLKKCISLTTGKYCMIMGNDD
LLADGALSKIVKVLKANPEIVLATRAYGWFKENPNELCDTVRHLTDDTLFQPGADAIK
FFFRRVGVISGFIVNAEKAKKLSSDLFDGRLYYQMYLAGMLMAEGQGYYFSDVMTLSR
DTEAPDFGNAGTEKGVFTPGGYKPEGRIHMVEGLLLIAKYIEDTTKIDGVYAGIRKDL
ANYFYPYIRDQLDLPLYTYIKMINKFRKMGFSNEKLFYVHAFLGYVLKRRGYDALIKY
IRSKKGGTPRLGI"
gene 5114. .6313
/gene="wzy"
CDS 5114. .6313
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O antigen polymerase"
/protein_id="CAA43912.1"
/db_xref="GI:47010"
/db_xref="GOA:Q00474"
/db_xref="UniProt/Swiss-Prot:Q00474"
/translation="MLPFPPGAILRDVLNVFFVALVLVRFVIDRKKTYFPLVFTIFSW
SAVILWVIALTIFSPDKIQAIMGGRSYILFPAVFIALVILKVSYPQSLNIEKIVCYII
FLMFMVATISIIDVLMNGEFIKLLGYDEHYAGEQLNLINSYDGMVRATGGFSDALNFG
YMLTLGVLLCMECFSQGYKRLLMLIISFVLFIAICMSLTRGAILVAALIYALYIISNR
KMLFCGITLFVIIIPVLAISTNIFDNYTEILIGRFTDSSQASRGSTQGRIDMAINSLN
FLSEHPSGIGLGTQGSGNMLSVKDNRLNTDNYFFWIALETGIIGLIINIIYLASQFYS
STLLNRIYGSHCSNMHYRLYFLFGSIYFISAALSSAPSSSTFSIYYWTVLALIPFLKL
TNRRCTR"
CDS 6313. .7323
/note="wbaW
Referred to as orf17.9 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="second mannosyl transferase"
/protein_id="CAA43913.1"
/db_xref="GI:47011"
/db_xref="GOA:Q00481"
/db_xref="UniProt/TrEMBL:Q00481"
/translation="MNNKKVLMDISWSNKGGIGRFTDEISKLLCDISKEELYRKCASP
LAPLGLAVNIFLRKKTDVVFLPGYIPPLFCSKKFIITIHDLNHLDLNDNSSLFKRLFY
NFIIKRGCRKAYKIFTVSNFSKERIVAWSGVNPNKIVTVYNGVSSLFNADVKPLNLGY
KYLLCVGNRKTHKNEKCVISAFAKADIDPSIKLVFTGNPCNDLEKLIIQHGLSERVKF
FGFVSEKDLPSLYKGSLGLVFPSLYEGFGLPVVEGMACGIPVLTSLTSSLPEVAGDAA
ILVDPLSEDAITKGISRLINDSELRKHLIQKGLLRAKRFNWQNVVSEIEMVLTEACDG
NK"
CDS 7310. .8467
/note="wbaZ
Referred to as orf18.9 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="first mannosyl transferase"
/protein_id="CAA43914.1"
/db_xref="GI:47012"
/db_xref="GOA:Q00483"
/db_xref="UniProt/TrEMBL:Q00483"
/translation="MEINEIKISLVHEWLLSYAGSEQVSSAILHVFPEAKLYSVVDFL
TDEQRRHFLGKYATTTFIQNLPKAKKFYQKYLPLMPLAIEQLDLSDANIIISSAHSVA
KGVISGPDQLHISYVHSPIRYAWDLQHQYLNESNLNKGIKGWLAKWLLHKIRIWDSRT
ANGVDHFIANSQYIARRIKKVYRREASVIYPPVDVDNFEVKNEKQDYYFTASRMVPYK
RIDLIVEAFSKMPEKKLVVIGDGPEMKKIKSKATDNIKLLGYQSFPVLKEYMQSARAF
VFAAEEDFGIIPVEAQACGTPVIAFGKGGALETVRPLGVEEEPTGIFFKEQNIASLHEA
VSEFEKNASFFTSQACRKNAEKFSRSRFEQEFKNFVNEKWNLFKTEQIIKR"
gene 8482. .9903
/gene="manC"
CDS 8482. .9903
/gene="manC"
/codon_start=1
/transl_table=11
/product="mannose-1-phosphate guanyltransferase"

```
                     /protein_id="CAA43915.1"
                     /db_xref="GI:47013"
                     /db_xref="GOA:Q00473"
                     /db_xref="UniProt/Swiss-Prot:Q00473"
                     /translation="MSKLIPVIMAGGIGSRLWPLSREEHPKQFLSVDGELSMLQNTIK
                     RLTPLLAGEPLVICNDSHRFLVAEQLRAINKLANNIILEPVGRNTAPAIALAAFCSLQ
                     NVVDEDPLLLVLAADHVIRDEKVFLKAINHAEFFATQGKLVTFGIVPTQAETGYGYIC
                     RGEAIGEDAFSVAEFVEKPDFDTARHYVESEKYYWNSGMFLFRASSYLQELKDLSPDI
                     YQACENAVGSINPDLDFIRIDKEAFAMCPSDSIDYAVMEHTRHAVVVPMNAGWSDVGS
                     WSSLWDISKKDPQRNVLHGDIFAYNSKDNYIYSEKSFISTIGVNNLVIVQTADALLVS
                     DKDSVQDVKKVVDYLKANNRNEHKKHLEVFRPWGKFSVIHSGDNYLVKRITVKPGAKF
                     AAQMHLHRAEHWIVVSGTACITKGEEIFTISENESTFIPANTVHTLKNPATIPLELIE
                     IQSGTYLAEDDIIRLEKHSGYLE"
     gene            9911. .11347
                     /gene="manB"

  Query Match           100.0%;  Score 1011;  DB 1;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 6.3e-160;
  Matches 1011;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6313 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 6372

Qy         61 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6373 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 6432

Qy        121 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6433 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 6492

Qy        181 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6493 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 6552

Qy        241 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6553 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 6612

Qy        301 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6613 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 6672

Qy        361 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6673 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 6732

Qy        421 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6733 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 6792

Qy        481 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6793 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 6852

Qy        541 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6853 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 6912

Qy        601 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6913 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 6972

Qy        661 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6973 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 7032

Qy        721 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7033 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 7092

Qy        781 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7093 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 7152

Qy        841 CCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7153 CCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTT 7212

Qy        901 CGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7213 CGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTG 7272

Qy        961 GTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGATGGAAATAAATGA 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7273 GTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGATGGAAATAAATGA 7323


RESULT 2
BD085556
LOCUS       BD085556              12441 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085556
VERSION     BD085556.1  GI:22631166
KEYWORDS    JP 2001522248-A/3.
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 12441)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 3 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Salmonella enterica
            PN   JP 2001522248-A/3
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .12441
            FT                   /organism='Salmonella enterica'.
FEATURES             Location/Qualifiers
     source          1. .12441
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match           100.0%;  Score 1011;  DB 6;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 6.3e-160;
  Matches 1011;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6313 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 6372

Qy         61 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6373 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 6432

Qy        121 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6433 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 6492

Qy        181 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6493 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 6552

Qy        241 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 300
```

```
Db       6553 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 6612

Qy        301 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6613 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 6672

Qy        361 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6673 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 6732

Qy        421 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6733 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 6792

Qy        481 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6793 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 6852

Qy        541 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6853 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 6912

Qy        601 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6913 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 6972

Qy        661 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6973 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 7032

Qy        721 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7033 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 7092

Qy        781 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7093 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 7152

Qy        841 CCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7153 CCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTT 7212

Qy        901 CGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7213 CGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTG 7272

Qy        961 GTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGATGGAAATAAATGA 1011
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7273 GTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGATGGAAATAAATGA 7323
```

```
RESULT 3
AX029263
LOCUS       AX029263               12446 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 3 from Patent WO9850531.
ACCESSION   AX029263
VERSION     AX029263.1  GI:10190114
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 3 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .12446
                     /organism="Salmonella enterica"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:28901"
ORIGIN
```

```
  Query Match           100.0%;  Score 1011;  DB 6;  Length 12446;
  Best Local Similarity  100.0%;  Pred. No. 6.3e-160;
  Matches 1011;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6313 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 6372

Qy         61 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6373 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 6432

Qy        121 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6433 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 6492

Qy        181 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6493 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 6552

Qy        241 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6553 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 6612

Qy        301 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6613 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 6672

Qy        361 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6673 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 6732

Qy        421 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6733 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 6792

Qy        481 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6793 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 6852

Qy        541 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6853 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 6912

Qy        601 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6913 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 6972

Qy        661 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6973 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 7032

Qy        721 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7033 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 7092

Qy        781 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7093 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 7152

Qy        841 CCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7153 CCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTT 7212

Qy        901 CGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7213 CGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTG 7272

Qy        961 GTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGATGGAAATAAATGA 1011
              |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       7273 GTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGATGGAAATAAATGA 7323


RESULT 4
AR384258
LOCUS       AR384258               1029 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 987 from patent US 6610836.
ACCESSION   AR384258
VERSION     AR384258.1  GI:40093992
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1029)
  AUTHORS   Breton,G.L. and Osborne,M.
  TITLE     Nucleic acid amino acid sequences relating to Klebsiella pneumoniae
            for diagnostics and therapeutics
  JOURNAL   Patent: US 6610836-A 987 26-AUG-2003;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1029
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match            26.8%;  Score 270.6;  DB 6;  Length 1029;
  Best Local Similarity  56.2%;  Pred. No. 1.2e-35;
  Matches  554;  Conservative    0;  Mismatches  419;  Indels   12;  Gaps    2;

Qy          1 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 60
              |||   || || ||  | ||  | |||| | |||||     | |    || || ||  |
Db         43 ATGGGAAAAAAGAAAATATTATTTGATACTCGTTGGATAGGTGAGCATGGTATAGGTAGA 102

Qy         61 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 120
              ||| || | ||| ||    |  |  ||    ||||    |     |||||    | |||
Db        103 TTTGCTAAGGAAGTTAGAGACTCAAAATTAACTATACAGGACTTAAAACTACCATGGAAA 162

Qy        121 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 180
                | |   |||||| |   || ||   ||  |  |  || | || |       ||||
Db        163 CCTACAAATCCGCTTGATACAATATTATTGACTATTTATTTAATTTTT------AAAAAA 216

Qy        181 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 240
              ||| |   ||||   ||||| || |   | ||  ||||||     |||   |  |||||
Db        217 GATATCTATTTTTCACCTGGATACAACGCGCCTTTTTTTTTCCTTAAACGATCTATAATC 276

Qy        241 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 300
              || || |||||||| |||||| | ||| |   | ||||    ||  || |||| ||  |
Db        277 ACGATTCATGATCTTAATCATATAGATATTCCTTATAACAGTTCCTTTCTTAAAAGAATT 336

Qy        301 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 360
              | |||||| ||| |  | ||  | | ||| |||||||| |     ||| | || || ||
Db        337 TATTATAAGTTTGTGCTTAAAAGAGCTTGCAGAAAAGCGTTATGTATACTAACCGTATCT 396

Qy        361 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 420
              | ||||| ||  | || ||    | |||||| |||  |||| |  || || |  |
Db        397 GAATTTTCTAAGAAGAGGATTTGTGAATGGTCGGGTCGAAACGCGGATGAAGTTTTTGTA 456

Qy        421 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 480
              |||   ||||| || || |     |||  |    ||     | ||   ||   ||
Db        457 GTAGGAAATGGAGTGTCCAAAGAGTTTCTTATAAATCACCCAACACCCAAGAAAG----- 511

Qy        481 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 540
              || |  |  ||  || || ||||| ||||  ||||| |||||   |  ||   ||
Db        512 -AACAAATATTAATGGTTGGTAACAGGAAAAAACATAAAAATGAATTGAATGCACTAATA 570

Qy        541 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 600
              || |||   ||  || ||||     ||||  | || || |||||||| |||  |  |
Db        571 GCTTTTTATAAGTCAAATATACCCTCATCGGTTAAGCTAGTTTTTACAGGTTCTATTACC 630

Qy        601 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 660
              ||||| ||||| | | | ||     |  |  | || ||  | |||   ||  ||||
Db        631 AATGAATTAGATATAATTATCGATCGTAACAGGCTTAGGGATAGGGTATCTTTTCTTGGT 690

Qy        661 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 720
              || ||  | |   |  | |  ||||||||  |  || |||   ||  | || |||
Db        691 ACAGTATCAAATACTAAACTTGCCGAGTTATATAGAGAATCCTTATTTTTGCTATTTCCT 750

Qy        721 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 780
              || |||||||||||||| ||  | ||  |||||||   |||| || || | ||| ||
Db        751 TCGTTATATGAAGGTTTCGGTCTCCCGATAGTGGAGTCAATGGCTTGCGGAACTCCAGTT 810

Qy        781 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 840
              |||  |||   |   |||| || || || || |||||||  || || ||| ||||  ||
Db        811 ATAATCTCTAATAGCACATCTTTACCTGAAGTTGCTGGAGGGGCTGCTATTATTGTAAAC 870

Qy        841 CCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTT 900
              ||   |   | ||| | ||| |   |   ||||   |  |||  | ||||| | ||||||
Db        871 CCAGATGATGTAGAGGATATAAAAGATAAAATTGAAAAATTATATCATGATGCAGAACTT 930

Qy        901 CGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTG 960
              |  ||  |  | ||| ||||||| |||||         | || |  |  ||| |||
Db        931 CAAAATAAACTTATCGAAAAGGGTCTTTTAAATATTGAAAGATATACCTGGGAAATGACT 990

Qy        961 GTTAGTGAGATTGAAATGGTACTGA 985
                | | |  ||||||   || |||
Db        991 CGCAATAAATTTGAAACAATAATGA 1015


RESULT 5
BA000016_25
WPCOMMENT
  Sequence split into 31 fragments  LOCUS BA000016 Accession BA000016
     Fragment Name      Begin      End
     BA000016_00            1    110000
     BA000016_01       100001    210000
     BA000016_02       200001    310000
     BA000016_03       300001    410000
     BA000016_04       400001    510000
     BA000016_05       500001    610000
     BA000016_06       600001    710000
     BA000016_07       700001    810000
     BA000016_08       800001    910000
     BA000016_09       900001   1010000
     BA000016_10      1000001   1110000
     BA000016_11      1100001   1210000
     BA000016_12      1200001   1310000
     BA000016_13      1300001   1410000
     BA000016_14      1400001   1510000
     BA000016_15      1500001   1610000
     BA000016_16      1600001   1710000
     BA000016_17      1700001   1810000
     BA000016_18      1800001   1910000
     BA000016_19      1900001   2010000
     BA000016_20      2000001   2110000
     BA000016_21      2100001   2210000
     BA000016_22      2200001   2310000
     BA000016_23      2300001   2410000
     BA000016_24      2400001   2510000
     BA000016_25      2500001   2610000
     BA000016_26      2600001   2710000
     BA000016_27      2700001   2810000
     BA000016_28      2800001   2910000
     BA000016_29      2900001   3010000
     BA000016_30      3000001   3031430
Continuation (26 of 31) of BA000016 from base 2500001 (BA000016 Clostridium perfringens s

  Query Match           11.3%;  Score 114.4;  DB 1;  Length 110000;
  Best Local Similarity  56.7%;  Pred. No. 3.1e-10;
  Matches  211;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy        585 TACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACG 644
              ||  ||   | | | | |   ||    ||||  | ||    |      ||||  |  |
Db      41846 TAAGGGACCTTCATATGAAATTTACAGAAAAAAAGTAGATGAGTTAAATTTATCTTCATC 41905

Qy        645 TGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTT 704
              ||| |  ||   ||| ||  |  ||  ||| ||| | ||    || ||||  |
```

```
Db      41906 TGTTATTTTTACTGGATTTATTCCTATAAATGATATGCCTATATTCTATAGTGCTAGTAA 41965

Qy        705 AGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGC 764

Db      41966 AGCATTAGTTTATCCTTCTTTTTATGAAGGCTTTGGATTACCTCCTATAGAATGCATGGC 42025

Qy        765 CTGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGC 824

Db      42026 CTGCGGAACCCCTGTAATAGCATCTAATTTAACCTCAATGCCTGAGGTATGCCAAGATGC 42085

Qy        825 AGCGATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAAT 884

Db      42086 TGCTCTTTTAGTTGATCCTTATGATGTTGATGAAATAAAGGAAAATATACTGACCTTATT 42145

Qy        885 TAATGATTCTGAACTTCGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTT 944

Db      42146 AAATAACCATAAATTTTATTCCCTTATGATTTATAAAGGGCTAAGTCATTCAAGTAAATT 42205

Qy        945 CAATTGGCAAAA 956

Db      42206 TAATTGGAAAAA 42217


RESULT 6
AJ749949_08
WPCOMMENT
  Sequence split into 19 fragments  LOCUS AJ749949 Accession AJ749949
     Fragment Name      Begin     End
     AJ749949_00            1    110000
     AJ749949_01       100001    210000
     AJ749949_02       200001    310000
     AJ749949_03       300001    410000
     AJ749949_04       400001    510000
     AJ749949_05       500001    610000
     AJ749949_06       600001    710000
     AJ749949_07       700001    810000
     AJ749949_08       800001    910000
     AJ749949_09       900001   1010000
     AJ749949_10      1000001   1110000
     AJ749949_11      1100001   1210000
     AJ749949_12      1200001   1310000
     AJ749949_13      1300001   1410000
     AJ749949_14      1400001   1510000
     AJ749949_15      1500001   1610000
     AJ749949_16      1600001   1710000
     AJ749949_17      1700001   1810000
     AJ749949_18      1800001   1892819
Continuation (9 of 19) of AJ749949 from base 800001 (AJ749949 Francisella tularensis sub

  Query Match           11.2%;  Score 113.6;  DB 1;  Length 110000;
  Best Local Similarity  50.4%;  Pred. No. 4.3e-10;
  Matches  390;  Conservative    0;  Mismatches  369;  Indels   15;  Gaps    4;

Qy        226 AAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCT 285

Db      17883 AAACCAATAATTATAACTATCCATGATCTTATACCTATTATATCTAAATATGAGAGTAAC 17942

Qy        286 CTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATAT--- 342

Db      17943 TTTATAAAGAAACTCTATTTTGAAAAAATACTAAAGCCGCTTATTAAAAGAGAAAATGTC 18002

Qy        343 AAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAAC 402

Db      18003 AAAATTATGACAGTTTCAAACTTTTCAAAAAATCAAATATTAGAATGGTCAGGGATTGAT 18062

Qy        403 CCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAA 462

Db      18063 AGCAGTAAAGTTATAGTAGCTTATAATGGCATATCAACAGCTTTT------GATTATAAG 18116

Qy        463 CCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAAT 522

Db      18117 CCAACTACAGTTAAGCAAAAATTTTTCTTATATATAGGTAATAGGCGCCCGCATAAAAAT 18176

Qy        523 GAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTT 582

Db      18177 ATTGAGATAATATTTAAAGCATTTAAAAACTTTAAGTATTACGATGAGTATAACTTAGTA 18236

Qy        583 TTTACTGGTAATCCTTGTAATGA---TTTAGAAAAACTAATAATACAACATGGTTTAAGT 639

Db      18237 TTCTCAGGAGAGATGGATCATCAACTGCTAAACTATGCTCAGAAATATAATGTTTGCCTC 18296

Qy        640 GAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGA---TTTACCATCGTTATATAAG 696

Db      18297 TCAAAGATACAAACTATAGGATTAGATGCTAATGATGAGCTTTTAGCTAAATATTATCAA 18356

Qy        697 GGCTCGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAG 756

Db      18357 TCAGCAACAGCAACGATACTACCTTCTTTTGAAGAGGGGTTTGGTTTACCTTTAGTTGAG 18416

Qy        757 GGCATGGCCTGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCT 816

Db      18417 TCTATGGCTTGTGGTACTCCTGTGCTAGCTTCTAAAATAGAAGTTTTACAGGAAATAGCA 18476

Qy        817 GGAGATGCAGCGATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCG 876

Db      18477 GCACAAGCGGGGTTGTATTTTGATCCTTATTCTGCTGATGAGTTATATAGTAATATTCAA 18536

Qy        877 AGGTTAATTAATGATTCTGAACTTCGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCA 936

Db      18537 AAACTGATTGATGACAAAGAGTTATATAACCAAAAGGTAGAGCTAAGTTTAGTAAGAGCA 18596

Qy        937 AAGAGGTTCAATTGGCAAAACGTGGTTAGTGAGATTGAAATGGTACTGACAGAG 990

Db      18597 AAACAGTTTAACTGGGATCAAACAGCAAATATTGTCCAAGATGAGCTTAAAAAG 18650


RESULT 7
AE007804/c
LOCUS       AE007804               12927 bp    DNA     linear   BCT 01-JUN-2004
DEFINITION  Clostridium acetobutylicum ATCC 824 section 292 of 356 of the
            complete genome.
ACCESSION   AE007804 AE001437
VERSION     AE007804.1  GI:15026130
KEYWORDS    .
SOURCE      Clostridium acetobutylicum ATCC 824
  ORGANISM  Clostridium acetobutylicum ATCC 824
            Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 12927)
  AUTHORS   Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
            Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
            Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
            Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
  TITLE     Genome sequence and comparative analysis of the solvent-producing
            bacterium Clostridium acetobutylicum
  JOURNAL   J. Bacteriol. 183 (16), 4823-4838 (2001)
   PUBMED   11466286
REFERENCE   2  (bases 1 to 12927)
  AUTHORS   Childress,D., Zeng,Q. and Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2001) GTC Sequencing Center Production,
            Finishing, and Bioinformatics Teams, Genome Therapeutics Corp., 100
            Beaver Street, Waltham, MA 02453-8443, USA
FEATURES             Location/Qualifiers
     source          1. .12927
                     /organism="Clostridium acetobutylicum ATCC 824"
                     /mol_type="genomic DNA"
                     /strain="ATCC 824"
                     /db_xref="ATCC:824"
                     /db_xref="taxon:272562"
     gene            complement(75. .1193)
                     /gene="CAC3068"
     CDS             complement(75. .1193)
                     /gene="CAC3068"
                     /codon_start=1
                     /transl_table=11
```

/product="Glycosyltransferase"
/protein_id="AAK81008.1"
/db_xref="GI:15026131"
/translation="MKVCMISGALPPIKCGVGDYSDILCSHLNKNSNIDLSVITSYDA
NKSYDYKIYNEIKSWNITELGRVLKLIKGISPDIVHIQYPTVSYKKGMMINILPYFLH
KKYKVITTLHEYTDSSILGKIRTNVGVLPSDLVIVVDDRYKKDLEKKKLFNKVRIEKI
NIGSNIPKSQKTYGEILSLKNKLLFHDKKKIIGYFGFVNNKKGIENILLAVKCLKDSG
YKDVQFNIMAELDASNAYHNKILKMIIDLGLNKDVLITGYLRKQVVGDYVKACDVMVL
PFLDGLSSKNGSFICALQEKRSIITTKPQIDNVYNKYKDERYVYFVDRVNNYKEIYNG
LIDLLFQNKFSTVGNNLNDFSWESIADEHYKVYKLLEN"

gene complement(1190. .2089)
/gene="CAC3069"

CDS complement(1190. .2089)
/gene="CAC3069"
/codon_start=1
/transl_table=11
/product="Predicted glycosyltransferase"
/protein_id="AAK81009.1"
/db_xref="GI:15026132"
/translation="MKSIYIILVNYNNFQDTIECVKSLRKIKNYHNYKVIIVDNDSKN
DSVSKIKSFLQDCLLIESKKNLGFAGGNNLGIKYAMKHNADYVLLLNNDTVVENDFLD
KLLEVHENYGNVGIVGGKIYYYYDKNRIWYAGGKINKFTGKTKHIGVNEIDEGQYNNI
SITGYVTGCMMLISKGAIEKCGVMDEKYFLYYEETDWNVRIKNAGFKIMYTPLSKIYH
KVSSTTGKSDLIMSYYYDRNRYYFVFKNFNWVCKLSVTLYILVRLYAKLFKAIAINDV
KRKHLIEYTIISIKNNIMGEIRV"

gene complement(2192. .3304)
/gene="CAC3070"

CDS complement(2192. .3304)
/gene="CAC3070"
/codon_start=1
/transl_table=11
/product="Glycosyltransferase"
/protein_id="AAK81010.1"
/db_xref="GI:15026133"
/translation="MTQNLRVIIDARMVNKHLHGIARYTYEIIKHMSYEDNVEMILLV
NDLEEASKIFGQFNNIKFKRMRSKFLSIFEQIELPMIVNKYKNMAIFHSPSFVASPFI
KCKMIMTIHDLNHLKFPQYYTPFHKYYYKYIVKKCALKSKKILTVSEFSKEEIFKWIK
CDRDKIKVTYNGIDPKFHVISDAEKLKKIKEKYKLPDRFILYIGNLKPHKNVETLVKA
MRFIDGKVKLVINGTINKSISEILDKFNLLEKVKFIGYVNEDELPLLYNLASVFVFPS
LYEGFGLPPLEAAVCGCKVIISDISSLSEIMSDYGVKVNPLDSKGMAREIKKILENSS
IDICNKEKETFVQKYSWNKTSSKTLEIYKEINKEFL"

gene complement(3304. .4422)
/gene="CAC3071"

CDS complement(3304. .4422)
/gene="CAC3071"
/codon_start=1
/transl_table=11
/product="Glycosyltransferase"
/protein_id="AAK81011.1"
/db_xref="GI:15026134"
/translation="MKVALVHDWLTNMGGAERVIINLKEIYKEAPIYTSLYNPDKLDD
ELRNIDVRTSFLQKPKNIKRGHQKLLPFMPMAFESFDLNEYDIVLSSCSSCAKGVVTN
PNTMHVCYCHTPMRYAWEYYYEYADEWGLNKLKGKLLKYVMNYIRLWDGVSSKRVDYY
IANSQNVAKRIWKHYKRESVVIHPPVRCKLFDISDVNEEYFLIVSRLVPYKRIDLAVE
VFNELGLPLVVIGGGKDLDRLKNMAKDNIKFLGRQSDEVIKEYYSKCRAFIFPGEEDF
GITPLEAQASGRPVIAYGKGGALETVVSGVTGVFFKEQTVDSLKSAVNEFEKINFDKS
VIRRHAEEFDEEIFKKKIREYLSRSYDEFKKLKKWKVI"

gene complement(4444. .5496)
/gene="CAC3072"

CDS complement(4444. .5496)
/gene="CAC3072"
/codon_start=1
/transl_table=11
/product="Mannose-1-phosphate guanylyltransferase"
/protein_id="AAK81012.1"
/db_xref="GI:15026135"
/translation="MLCALIMAGGKGERFWPLSTDEKPKQFLKLLSDKTMIQMTVERL
ESLIDIEKIFVVTGERYVDILREQLPRLPKRNIIVEPVGRNTAPCIALSAFVINKYYK
DASIVVLPSDHLIVNEDKLLDAIRSAEEFIEKNDNSIVTLGMEPTRAETGYGYIKSGE
VKEEISGFKIKKVDEFVEKPNKEKAEEYVKAGNFLWNGGMFIWKCSTIFKLTEKYLNK
TYNILRSISEANGDEFNVLLKREYVKVDNISVDYGIMEKADSIYVIPCDFGWDDIGTW
HAVERYRDKDEDNNVCVGDIKSIDSKNNILVGKKKPIVVVGLSDVFVVESDDIIFVGK
KDDIERIKEIKKKVIK"

gene complement(5508. .6176)
/gene="CAC3073"

CDS complement(5508. .6176)
/gene="CAC3073"
/codon_start=1
/transl_table=11
/product="Sugar transferase involved in lipopolysaccharide
synthesis"
/protein_id="AAK81013.1"
/db_xref="GI:15026136"
/translation="MQNLELEDKKVVFMPEESLDKGIGYFIIKRIIDVFGSLVGIIFL
SPVMLIVIVAIKLDSKGPIIFSQKRVGQNGRIFNMYKFRSMVENAERLLGKLQDKNEM
SGPMFKMSDDPRITKVGRFIRKTSLDELPQLFNVLRGHMSLVGPRPNLPNEVKEFNEF
QKRKFMVKPGLTCYWQVMGRSSIGFEEWIELDLKYIKERSTFVDLKLIFRTFFLLFGD
KNAM"

gene 6470. .6955
/gene="CAC3074"

CDS 6470. .6955
/gene="CAC3074"
/codon_start=1
/transl_table=11
/product="Uncharacterized conserved protein"
/protein_id="AAK81014.1"
/db_xref="GI:15026137"
/translation="MKKIFLILLCILCFSFIFYNSSQNANVSNDRSYRITNDLRNFYR
KLKGENQRSYTKLPTNPRDEKINIFIRKNAHAFEYFLLACVVTIIIFSLGLKGRYAVV
YIWFICLFYAVLDEFHQLYVPGRTSLVSDVLVDFLGANIGMWISYFVYYVILKKFIRK
K"

gene complement(7036. .8103)
/gene="CAC3075"
/note="buk"

CDS complement(7036. .8103)
/gene="CAC3075"
/codon_start=1
/transl_table=11
/product="Butyrate kinase, BUK"
/protein_id="AAK81015.1"
/db_xref="GI:15026138"
/translation="MYRLLIINPGSTSTKIGIYDDEKEIFEKTLRHSAEEIEKYNTIF
DQFQFRKNVILDALKEANIEVSSLNAVVGRGGLLKPIVSGTYAVNQKMLEDLKVGVQG
QHASNLGGIIANEIAKEINVPAYIVDPVVVDELDEVSRISGMADIPRKSIFHALNQKA
VARRYAKEVGKKYEDLNLIVVHMGGGTSVGTHKDGRVIEVNNTLDGEGPFSPERSGGV
PIGDLVRLCFSNKYTYEEVMKKINGKGGVVSYLNTIDFKAVVDKALEGDKKCALIYEA
FTFQVAKEIGKCSTVLKGNVDAIILTGGIAYNEHVCNAIEDRVKFIAPVVRYGGEDEL
LALAEGGLRVLRGEEKAKEYK"

gene complement(8131. .9036)
/gene="CAC3076"
/note="ptb"

CDS complement(8131. .9036)
/gene="CAC3076"
/codon_start=1
/transl_table=11
/product="Phosphate butyryltransferase"
/protein_id="AAK81016.1"
/db_xref="GI:15026139"
/translation="MIKSFNEIIMKVKSKEMKKVAVAVAQDEPVLEAVRDAKKNGIAD
AILVGDHDEIVSIALKIGMDVNDFEIVNEPNVKKAALKAVELVSTGKADMVMKGLVNT
ATFLRSVLNKEVGLRTGKTMSHVAVFETEKFDRLLFLTDVAFNTYPELKEKIDIVNNS
VKVAHAIGIENPKVAPICAVEVINPKMPSTLDAAMLSKMSDRGQIKGCVVDGPLALDI
ALSEEAAHHKGVTGEVAGKADIFLMPNIETGNVMYKTLTYTTDSKNGGILVGTSAPVV
LTSRADSHETKMNSIALAALVAGNK"

gene complement(9145. .10767)
/gene="CAC3077"

CDS complement(9145. .10767)
/gene="CAC3077"
/codon_start=1
/transl_table=11
/product="Uncharacterized FAD-dependent dehydrogenase"
/protein_id="AAK81017.1"
/db_xref="GI:15026140"
/translation="MIRINNIILDIDEDLENLKNKAAKKLKVSKESFESFKILRESID
ARKKDVIKFNYSVEVKCKNESKIISKIHGKDAILEKNKEEERFQFGNEKLKHRPIVIG
TGPCGMFAALTLAENGYAPLVIERGENVEKRTETVDNFWKTGVLNTESNVQFGEGGAG

```
                     TFSDGKLTTRIKDPRCTYVLEEFIKSGAPEEIGYVGKPHIGTDILKKVVKNIRERIIS
                     LGGEVKFNSKLEDIIIKNDSIEAVVVNGAEIPCECLVLAIGHSSRDTYEMLHKKGIFM
                     DPKPFAIGVRIEHPQDFINESQYGKYKNHPRLKAADYKLTHTTKDKRGVYSFCMCPGG
                     EVVASSSEDERLVCNGMSYYKRDLENANAALVVTVNSEDFKASKELYNKDISDNSPLI
                     GMEFQRYYEHLAYLAGGGNYNAPIQLVGDFMKDTVSSKLGSVKPSYTPGYKFASLSEC
                     LPPYVIAALKEGLVNFDKKITGYMLSDAVMTGIETRTSAPLKITRNESLESISLKGLY
                     PSGEGAGFAGGIISAAVDGVKSAESIMRKYKI"
     gene            complement(10783. .11994)
                     /gene="CAC3078"
     CDS             complement(10783. .11994)
                     /gene="CAC3078"
                     /codon_start=1
                     /transl_table=11
                     /product="Uncharacterized secreted protein, YBBR Bacillus
                     subtilis homolog"
                     /protein_id="AAK81018.1"

  Query Match             11.1%;  Score 112.6;  DB 1;  Length 12927;
  Best Local Similarity   47.8%;  Pred. No. 1.3e-09;
  Matches  380;  Conservative    0;  Mismatches  394;  Indels   21;  Gaps    1;

Qy        118 AAATGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAA 177
              || | |   |||    | |  | | | |  ||  | | | | ||  |  |   || ||
Db       3103 AAGTTTTTATCTATTTTTGAACAAATTGAATTGCCTATGATTGTTAATAAGTATAAAAAT 3044

Qy        178 ACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATA 237
              |  | | |  ||  |   |||   | | || ||   |  |||      |   ||| | ||
Db       3043 ATGGCTATATTTCATTCGCCTTCTTTTGTTGCAAGCCCGTTTATAAAGTGTAAGATGATT 2984

Qy        238 ATAACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGG 297
              || |||||||||||| | ||||||||  | ||   | |  |||   | | |||| | | |
Db       2983 ATGACAATACATGATTTGAATCATCTAAAATTTCCTCAATATTATACACCTTTTCATAAG 2924

Qy        298 TTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTT 357
              |  | ||| || |  ||| |||||     ||       |||     | ||||||| || |||
Db       2923 TATTATTACAAATACATAGTAAAGAAATGTGCGTTGAAAAGTAAAAAAATATTAACTGTT 2864

Qy        358 TCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTC 417
              ||  | |||||  || ||   |||  |    |||  |        |      |||||||
Db       2863 TCTGAATTTTCTAAGGAGGAAAATTTTTAAGTGGATAAAGTGTGATAGGGATAAAATTAAA 2804

Qy        418 ACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGC 477
              |||||||  |   |  |  |||| |||        |  ||  ||| |||
Db       2803 GTCACGTATAATGGAATTGATCCTAAATTTCATGTGATAAGTGATGCAGAGAAATTAAAA 2744

Qy        478 TATA--------------------AATATTTGCTATGTGTAGGAAACAGAAAAACTCAT 516
              |  ||                  | |||     |  | ||   | | | ||      |
Db       2743 AAAATAAAAGAAAAATATAAGTTGCCAGATAGATTTATTTTATATATCGGTAACTTAAAG 2684

Qy        517 AAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAA 576
                || | ||   ||  | ||  |      || |     |  ||||||       | | |||
Db       2683 CCACATAAAAATGTTGAAACATTAGTAAAAGCTATGAGATTTATTGACGGCAAAGTTAAA 2624

Qy        577 CTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTA 636
              | ||  |||  || |         |||   | ||    |  |  ||   ||  |  |||
Db       2623 TTGGTAATTAACGGAACAATAAATAAATCTATAAGTGAGATTTTAGACAAATTTAATTTG 2564

Qy        637 AGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAG 696
              ||    || || ||  | || |  ||   |||  | || ||||||    | |||||
Db       2563 CTCGAGAAAGTTAAATTTATAGGTTATGTTAATGAGGATGAATTACCACTACTGTATAAT 2504

Qy        697 GGCTCGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAG 756
              |    |  || || || || || ||||||||||| |||||  ||||    | | |||
Db       2503 TTGGCTAGCGTTTTTGTGTTTCCATCATTATATGAAGGATTTGGGCTACCACCATTAGAG 2444

Qy        757 GGCATGGCCTGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCT 816
              |      |  |||||    |   ||  | |  |||    |  || ||||||   ||  |
Db       2443 GCTGCTGTGTGTGGATGTAAAGTTATTATCTCTGATATATCGTCATTGAGTGAAATTATG 2384

Qy        817 GGAGATGCAGCGATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCG 876
              | |||   |   ||   ||  | |||||      | | |  ||  | | ||  |||
Db       2383 AGTGATTATGGAGTTAAGGTGAATCCTCTAGATTCAAAAGGAATGGCAAGAGAGATTAAA 2324
```

```
Qy         877 AGGTTAATTAATGAT 891
               |   || |  |  ||
Db        2323 AAAATATTAGAAAAT 2309


RESULT 8
BA000019_53
WPCOMMENT
  Sequence split into 65 fragments  LOCUS BA000019 Accession BA000019
     Fragment Name      Begin      End
     BA000019_00            1   110000
     BA000019_01       100001   210000
     BA000019_02       200001   310000
     BA000019_03       300001   410000
     BA000019_04       400001   510000
     BA000019_05       500001   610000
     BA000019_06       600001   710000
     BA000019_07       700001   810000
     BA000019_08       800001   910000
     BA000019_09       900001  1010000
     BA000019_10      1000001  1110000
     BA000019_11      1100001  1210000
     BA000019_12      1200001  1310000
     BA000019_13      1300001  1410000
     BA000019_14      1400001  1510000
     BA000019_15      1500001  1610000
     BA000019_16      1600001  1710000
     BA000019_17      1700001  1810000
     BA000019_18      1800001  1910000
     BA000019_19      1900001  2010000
     BA000019_20      2000001  2110000
     BA000019_21      2100001  2210000
     BA000019_22      2200001  2310000
     BA000019_23      2300001  2410000
     BA000019_24      2400001  2510000
     BA000019_25      2500001  2610000
     BA000019_26      2600001  2710000
     BA000019_27      2700001  2810000
     BA000019_28      2800001  2910000
     BA000019_29      2900001  3010000
     BA000019_30      3000001  3110000
     BA000019_31      3100001  3210000
     BA000019_32      3200001  3310000
     BA000019_33      3300001  3410000
     BA000019_34      3400001  3510000
     BA000019_35      3500001  3610000
     BA000019_36      3600001  3710000
     BA000019_37      3700001  3810000
     BA000019_38      3800001  3910000
     BA000019_39      3900001  4010000
     BA000019_40      4000001  4110000
     BA000019_41      4100001  4210000
     BA000019_42      4200001  4310000
     BA000019_43      4300001  4410000
     BA000019_44      4400001  4510000
     BA000019_45      4500001  4610000
     BA000019_46      4600001  4710000
     BA000019_47      4700001  4810000
     BA000019_48      4800001  4910000
     BA000019_49      4900001  5010000
     BA000019_50      5000001  5110000
     BA000019_51      5100001  5210000
     BA000019_52      5200001  5310000
     BA000019_53      5300001  5410000
     BA000019_54      5400001  5510000
     BA000019_55      5500001  5610000
     BA000019_56      5600001  5710000
     BA000019_57      5700001  5810000
     BA000019_58      5800001  5910000
     BA000019_59      5900001  6010000
     BA000019_60      6000001  6110000
     BA000019_61      6100001  6210000
     BA000019_62      6200001  6310000
```

```
     BA000019_63       6300001   6410000
     BA000019_64       6400001   6413771
Continuation (54 of 65) of BA000019 from base 5300001 (BA000019 Nostoc sp. PCC 7120 DNA,

  Query Match             11.0%;  Score 110.8;  DB 1;  Length 110000;
  Best Local Similarity   58.5%;  Pred. No. 1.3e-09;
  Matches  193;  Conservative    0;  Mismatches  137;  Indels    0;  Gaps    0;

Qy        626 AACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCAT 685
              ||| ||  ||| ||| ||||| |||||| |||  |  || ||| |  | || |||||
Db      81438 AACTAGGGATAACTGATCGTGTGAAGTTCCTTGATTATGTTTCTTATGATGAATTACCTA 81497

Qy        686 CGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTAC 745
                |   |||      || | | | | |||||||||   | | |  ||||| ||||| || |
Db      81498 TCATTATTAATCAAGCGATCGCACTTGTTTTCCCCAGTCTTTGGGAAGGCTTTGGTTTTC 81557

Qy        746 CTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGC 805
              | ||  | || | | ||||||| |||| ||||||  | || |||       || || || |
Db      81558 CAGTCTTAGAAGCCTTGGCCTGCGGTACTCCTGTCATTACCTCTAACCTCTCTTCCTTAC 81617

Qy        806 CAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTA 865
              | || ||||||||||||||||| || ||  || ||||       ||  |  || ||
Db      81618 CTGAAGTGGCTGGAGATGCAGCCATCCTCATCAACCCCTACAACACAGCAGAAATCACAG 81677

Qy        866 AAGGAATTTCGAGGTTAATTAATGATTCTGAACTTCGTAAGCATTTAATCCAAAAGGGGC 925
              | |  ||   ||   ||  |||||||    || | |     ||| |    |   | ||
Db      81678 ATGCCATGCAGACTATAGCTAATGATGTCAAATTGCAATCCCATCTGTCTCGCCAAGGAA 81737

Qy        926 TTTTGCGGGCAAAGAGGTTCAATTGGCAAA 955
              |     | || ||    |||| |||| |||
Db      81738 TCGCTAGAGCCAATCAATTCAGTTGGGAAA 81767


RESULT 9
AE000693
LOCUS       AE000693               14693 bp    DNA     linear   BCT 17-MAR-2003
DEFINITION  Aquifex aeolicus VF5 section 25 of 109 of the complete genome.
ACCESSION   AE000693 AE000657
VERSION     AE000693.1  GI:2983148
KEYWORDS    .
SOURCE      Aquifex aeolicus VF5
  ORGANISM  Aquifex aeolicus VF5
            Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
REFERENCE   1  (bases 1 to 14693)
  AUTHORS   Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L.,
            Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M.,
            Huber,R., Feldman,R.A., Short,J.M., Olsen,G.J. and Swanson,R.V.
  TITLE     The complete genome of the hyperthermophilic bacterium Aquifex
            aeolicus
  JOURNAL   Nature 392 (6674), 353-358 (1998)
   PUBMED   9537320
REFERENCE   2  (bases 1 to 14693)
  AUTHORS   Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L.,
            Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M.,
            Huber,R., Feldman,R.A., Short,J.M., Olson,G.J. and Swanson,R.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego,
            CA 92121
COMMENT     Putative indicates no similarity to known proteins
            Hypothetical indicates similarity to a protein of unknown function.
FEATURES             Location/Qualifiers
     source          1. .14693
                     /organism="Aquifex aeolicus VF5"
                     /mol_type="genomic DNA"
                     /strain="VF5"
                     /db_xref="taxon:224324"
     gene            complement(72. .1043)
                     /gene="trxB"
                     /note="synonym: aq_500"
     CDS             complement(72. .1043)
                     /gene="trxB"
                     /codon_start=1
                     /transl_table=11
                     /product="thioredoxin reductase"
                     /protein_id="AAC06756.1"
                     /db_xref="GI:2983154"
                     /translation="MAVSLMQQPDKVYDVIIIGAGPAGTTAAIYTARAGWKTLVLYRA
                     EADGALGVTQKIENYPGVPGPLSGYELLKIMREQAKSFGAEFVRGKVIATDLNSDPKK
                     VYTIDGREFRGKTIIVASGAMERANKFKGEEEFLGRGVSYCGVCDAAFFKDQPVAVIG
                     DDDYAIEEAEFIARFANKVFFVVPGSKIKAPPEVIEHFEKLPNVEILLRHRPIEIVGD
                     QVVKGIKLKDLEKKEEKLLEVNGVFIFLGGTKPSVDFLMGQVEMTEGDCIVVNEEMMT
                     SVPGVFAAGDVLCNEVKQAVVAAAMGCKAALAVDKFLSGKKKIVPQW"
     gene            complement(1186. .2685)
                     /gene="pmu"
                     /note="synonym: aq_501"
     CDS             complement(1186. .2685)
                     /gene="pmu"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoglucomutase/phosphomannomutase"
                     /protein_id="AAC06757.1"
                     /db_xref="GI:2983155"
                     /translation="MEQGYRTISRKTLRPLQKDIRKITFLCKAECSRKRIIQYSIMIK
                     FGTDGWRARIAEEFNFDNVRRVAKAHAEVLKEKGVKRVIVGYDWRFRSEDFAKAVYDV
                     FRSEGLEAKLVGSACTTPMVSFAVKYLGYENGVMITASHNPPAYNGYKIKESFGGSAT
                     PEFVKNVEEKVKDIESVEIKEFEPEYQEVRKKYVEKIREFFDMDLFKEKKTLVVHDSM
                     HGTSSGLLDFVLRDTKVSVINIRKDRDPLFGGHPPEPIEKYMEITKEKVRALGAELGI
                     ANDGDGDRIALVDDKGNYVNTQLIYAMLLLHLLENKGKRGKVVKTVSTTYLADRICKE
                     FGVPLIEVGVGFKNINEIILKEGGVIFGGEESGGYGIPEYLPERDGIFSALNILELLY
                     LKDKKVSEVVAEIFERFGEAYYKRVDLRVKEKIKGRIKELKENPPERIGEFKVREVKT
                     IDGVKFVFEDDSWLLMRPSGTEPLIRVYAETPTKEEVEKIIQKGLELLK"
     gene            complement(2577. .3038)
                     /gene="aq_503"
     CDS             complement(2577. .3038)
                     /gene="aq_503"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAC06760.1"
                     /db_xref="GI:2983158"
                     /translation="MGIGGDYVWKRKPDVFLGIRKDWDFYDFYVSAYYMTRKPQLLFS
                     VKAGRFLAGDKGVRIEVSRIVKGFEVGFWYTYSDTSDFTGPNKNYHDKGVFVRIPLRI
                     FNLRDVKQIASFSLAPWSRDIGQLAGRPFDLYRKIYEKLPFYVKQNAAEKE"
     gene            complement(3122. .3541)
                     /gene="otnA'"
                     /note="synonym: aq_504"
     CDS             complement(3122. .3541)
                     /gene="otnA'"
                     /codon_start=1
                     /transl_table=11
                     /product="polysaccharide biosynthesis protein (fragment)"
                     /protein_id="AAC06755.1"
                     /db_xref="GI:2983153"
                     /translation="MEDGDMIYVPRKPSAVLVFGEVQNPSAILYKPGLKVKDYIALSG
                     GFTRDADVDNVFIIKANGIAISQESGKSLIEWDSERKRFIWGWAYNDILNYKLEPGDA
                     IIVPTKIKIPTMWRPLIRDVIQIIYQSALTVYTITKL"
     gene            complement(3572. .5749)
                     /gene="otnA"
                     /note="synonym: aq_505"
     CDS             complement(3572. .5749)
                     /gene="otnA"
                     /codon_start=1
                     /transl_table=11
                     /product="polysaccharide biosynthesis protein"
                     /protein_id="AAC06754.1"
                     /db_xref="GI:2983152"
                     /translation="MQGRDTVRMWRDRKMKVFLRSLILIFVAAFLFVYAKDKPQPKKL
                     PKYKNLISKIEKEFSSRNSTPLRQFGYEFFMTPLDVSKLVLPVSENYKLGPGDEVIIY
                     IVGLPSNQEIPPVIKTIVDREGNISIPNLGVFNVWGLTLKETEELLRQELGLDLKVTV
                     GEIRTFVIYVAGEVNRPGTHVATAVNTVIDALSIAGGVKKTGSLRNIILTRKTPSGLK
                     TYNIDLYDVFIKGKPIDLTLRDGDTILVKPIGPTVAIAGEVKRPAIYEIKGGETIKDL
                     IELAGGLLPSSYQYKVILQRYENGKLRIIEDSLTNKAFMNENLRDGDLVLIKKSADIP
                     ENSYSIEGYVTLPGVYEYKEGLKLSDILNPDLFFKDTNLKFAVLVRQYPKGSIPKYIA
                     FSPEKVLTGKEDMELKPYDRLIFFKLGQSVGKKINLNLVKNYVLVSGYYKYTGPVACS
                     DDCYISEILQKDLILKDTNLNYAEIERFNPKTLQREEIIKFKPIEVLNGSKDIKLQSW
                     DRIVFYPEFVYEPVRISGEVERPLWVTYKPGLTLEEAISQVKFKYPLGELKVIVKQDK
```

```
                KVEVFYLNDLRKHPSPLSFVKLKPGTEIVVKRVKEDETVTKVYVYGQVWKPGVYKIGE
                RTTLYDVLKAAGGFKPMAYPKGIVILRESIAQMQRERLNKAMVFLKAQLEKEIAGYMQ
                AELPKEQKEAYKEAFAAKRRLLQEMEKVQVTGRLVGLNIPPDI"
gene            complement(5691. .6992)
                /gene="aq_506"
CDS             complement(5691. .6992)
                /gene="aq_506"
                /codon_start=1
                /transl_table=11
                /product="putative protein"
                /protein_id="AAC06758.1"
                /db_xref="GI:2983156"
                /translation="MFLRLKKSSIYFLRSSIVSIFAIVSLKFIRTLNLASYKMISSME
                EKSIDRKSEKLGYPMYEEDEIDLYELWVVLKKRFRVILLTLVIFIAGAVAYVFIAPPV
                YKTEAVIFPIGGEKTNLPSILEGLPVPLPSVADKVTVESVLKSRTLRERVIEKLNLMP
                VLFKDLWDESRNTWRQDLEPEDIPTILDGAELLEDLISVSSNKKTGTITFSVEFPEDP
                HMAYKIAQVSLEEAYKILNEKSWSLAKKHRIFLENQIKEVLNKQKVLEEVYLKFIKGE
                IKEVPIVVDERFISKIFKESLNYENNDKFDIPGNLKQNLERNIPIAKGMNRNLPEYQF
                NYQKLQWQFEVLKNLLSTLFQQYQIAKANEVKESVSFQIIDPPYIPQKPAKPKKALIL
                AVSFVSGLFTGIFLAFFKEWLNNARKRYSENVEGQENESVS"
gene            6700. .7236
                /gene="aq_507"
CDS             6700. .7236
                /gene="aq_507"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAC06761.1"
                /db_xref="GI:2983159"
                /translation="MNTYATAPAIKITRVSSITLNLFLSTTQSSYRSISSSSYIGYPN
                FSLFRSILFSSILDIILYDAKLSVRINFKLTMAKIETIEDLRKYIEDFFKRRNKKVKV
                ILFGSRARGTHTSYSDVDIAILSDENIDNDLILLREEIENSLLPQKVDIVGFSKLEES
                FKEEILKEGKVWIDLTKS"
gene            8028. .9353
                /gene="aq_509"
CDS             8028. .9353
                /gene="aq_509"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAC06762.1"
                /db_xref="GI:2983160"
                /translation="MGIEEKLTTASKIKSLLNIREQEKVLDALIRSSFINILARGFGY
                LKHVTIAVVLGFNLQTDAFFMALSLIGIFLIFADVFDSIGVPNLVDARLKSQEEFKKL
                SGLLFTFTLILTFLTGFLAIGFYPFVLKIPVGFNMEALSYTQVAYFLLLPYLLFNFLF
                HHFGAVLRSVRRFTVYFVGEFLFSFFSFLFIFLGLIFYNKWLVLPLSISFAQIIATLY
                MFYVGREFIHINFFLNQTTKKILNHFFYLSALYGVLHLFIVIDRAFGSLLGEKSISAI
                HYGLIVASIPRGIIRLENIAITSLSESKGDIKKLNFYIKRILIVSGISLLFFFVFAEI
                IVKLLFGYGAFTQMDLILTVEATRFYSLSLPFMFLWPILYRVFQIKENLKPVFFIAIS
                GITVNGILNYFFVLKLNYGIKGIALGTLGAYIVICSLSYIILYYKYRRN"
gene            9354. .9824
                /gene="aq_510"
CDS             9354. .9824
                /gene="aq_510"
                /codon_start=1
                /transl_table=11
                /product="putative protein"
                /protein_id="AAC06759.1"
                /db_xref="GI:2983157"
                /translation="MVKSDYTRCPICETQKNVNQYIDTYISPFNNQEYKLYECSNCKL
                QWWEPLKIIPEFYENEVFDSYISFHEGIRSRIGKNHEAFFKYVPKNVKGKLLDIGCGD
                GVFLREAQKYGFEVWGIDFDKKSVETAKKNLGVKTIYAMSLEEFHKFAKDNNIR"
gene            9863. .10504
                /gene="murB2"
                /note="synonym: aq_511"
CDS             9863. .10504
                /gene="murB2"
                /codon_start=1
                /transl_table=11
                /product="UDP-N-acetoenolpyruvoylglucosamine reductase"
                /protein_id="AAC06751.1"
                /db_xref="GI:2983149"
                /translation="MEAFAGTPLKDLIRFSVKENIRNFYKLLGFPASVGGAVSMNAGA
                FGVEISDFLKEVYFVDWEGKLQKAKRDELSFSYRKSPFPELGIVYKVVFELGKSEENI
                LPKYEKIRRIRKEKQPINFPTSGSTFKNPKGNFAGKLLEEVSLKGFRLKNVGFSEKHA
                NFLVNYGGGTFSEVVDLINIAKERVYENFGIVLEEEVKLIESSGSDGWKVLGA"
gene            complement(10341. .11324)
                /gene="aq_512"
CDS             complement(10341. .11324)
                /gene="aq_512"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAC06763.1"
                /db_xref="GI:2983161"
                /translation="MEQGYRTISRKTLRPLQKDIRKITFLCKAECSRKRIIQYSIMIK
                FGTDGWRARIAEEFNFDNVRRVAKAHAEVLKEKGVKRVIVGYDWRFRSEDFAKAVYDV
                FRSEGLEAKLVGSACTTPMVSFAVKYLGYENGVMITASHNPPAYNGYKIKESFGGSAT
                PEFVKSVEEKVKDLESVEVREFEPEYQEVRKKYVEKIREFFDMDLFKEREALIVHDSL
```

```
Query Match             10.5%;  Score 106;  DB 1;  Length 14693;
Best Local Similarity   48.7%;  Pred. No. 1.6e-08;
Matches  348;  Conservative    0;  Mismatches  360;  Indels    6;  Gaps    2;

Qy       165 TCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTC 224
            |||   ||| |||  | || || ||||  |||||  |   | || ||| |||| | ||
Db     12671 TCTTGAAAATAAAGTTTATTTGTTTTTCTTTCCTCACATAAATCTACCTCTTTATATATC 12730

Qy       225 GAAAAAGTTCATAATAACAATACATGATCTAAATCATCTG---GATTTAAATGATAATTC 281
             ||||   |  ||||  | ||||||||||  | | |    || |    ||||
Db     12731 ACAAAATACTGTTGTAACCGTTCATGATCTAAGACCTTTCACAGAGTATTGGGATAGAGG 12790

Qy       282 CTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATA 341
               |   ||     |  || | ||  || |  |  | | |    ||  |  || |
Db     12791 AGAATTCAAAAAACTTATTATTAAAAAGCTTTTTTTTGATAGAGCAATTAAGTATTCAAA 12850

Qy       342 TAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAA 401
             |||  |   |  | || ||      | ||| |      || ||  ||  |      | |
Db     12851 GGCTATAATATCTATATCAAAAACAACTAAAAACGACTTACTAAAATTTTATTCCGAAGT 12910

Qy       402 CCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAA 461
              |  ||||| || ||     | ||| |     |  || |     ||  |     |  |
Db     12911 TCAAAATAAGATTGTGGTTATTTATGAAACATTTAGTATTAATAAAAAATCAACAGAGAC 12970

Qy       462 ACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAA 521
             |  |   ||| | |      | ||| |||| | ||||||  |     ||||  ||||| ||
Db     12971 AAAACAAAATATTGTTGGGGATTATCTGCTCTATGTAGGGCAGTTTAAAAAACATAAAAA 13030

Qy       522 TGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGT 581
             |    | | |  |    || ||| | |      | | | |       |    || || ||
Db     13031 TTTAGATAGATTGCTTAAAGCTTTTTCTATGATCAAAAATAAAATTAAATGCAAGCTTGT 13090

Qy       582 TTTTACTGGTAATCCTTGTAAT---GATTTAGAAAAACTAATAATACAACATGGTTTAAG 638
             |  | || ||       ||||   ||| | | ||   || | |   |    | | |
Db     13091 AATAGCAGGAAAAAAAGATAATATAGATATCTATAATTCTATCAAAAGCCTCAATCTTGG 13150

Qy       639 TGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGG 698
             ||| |||||   | | |          || ||    || ||   |  ||||||||
Db     13151 TGATCGTGTTGTGGTTATAAAAAATCCTTCCGATGGCGAGTTGATAAATTTATATAAAAA 13210

Qy       699 CTCGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGG 758
             |    |   ||  |||  |||||||||||||||||||||||| |||||    |  ||  |
Db     13211 TGCTAAATTTTTCATTTCTCCTTCTTTATATGAAGGTTTTGGTTTACCACCCCTTGAAAG 13270

Qy       759 CATGGCCTGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGG 818
             |    | ||  | | |||||| ||   |||    |   ||  |||   ||  |   |
Db     13271 TTTAAATTATGATGTCCCTGTAATATTGTCTGATATACCAGTATTTAGGGAAATATACGA 13330

Qy       819 AGATGCAGCGATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAAGGAAT 872
             ||||  ||| |||  | |  | ||| |||| |||||   |||| || ||   ||
Db     13331 AGATATAGCAATTTATTTTAATCCTTTTTCAGAAGAAAATATTGCTGAAAAGAT 13384


RESULT 10
CP000099_02/c
WPCOMMENT
```

Sequence split into 49 fragments LOCUS CP000099 Accession CP000099

Fragment Name	Begin	End
CP000099_00	1	110000
CP000099_01	100001	210000
CP000099_02	200001	310000
CP000099_03	300001	410000
CP000099_04	400001	510000
CP000099_05	500001	610000
CP000099_06	600001	710000
CP000099_07	700001	810000
CP000099_08	800001	910000
CP000099_09	900001	1010000
CP000099_10	1000001	1110000
CP000099_11	1100001	1210000
CP000099_12	1200001	1310000
CP000099_13	1300001	1410000
CP000099_14	1400001	1510000
CP000099_15	1500001	1610000
CP000099_16	1600001	1710000
CP000099_17	1700001	1810000
CP000099_18	1800001	1910000
CP000099_19	1900001	2010000
CP000099_20	2000001	2110000
CP000099_21	2100001	2210000
CP000099_22	2200001	2310000
CP000099_23	2300001	2410000
CP000099_24	2400001	2510000
CP000099_25	2500001	2610000
CP000099_26	2600001	2710000
CP000099_27	2700001	2810000
CP000099_28	2800001	2910000
CP000099_29	2900001	3010000
CP000099_30	3000001	3110000
CP000099_31	3100001	3210000
CP000099_32	3200001	3310000
CP000099_33	3300001	3410000
CP000099_34	3400001	3510000
CP000099_35	3500001	3610000
CP000099_36	3600001	3710000
CP000099_37	3700001	3810000
CP000099_38	3800001	3910000
CP000099_39	3900001	4010000
CP000099_40	4000001	4110000
CP000099_41	4100001	4210000
CP000099_42	4200001	4310000
CP000099_43	4300001	4410000
CP000099_44	4400001	4510000
CP000099_45	4500001	4610000
CP000099_46	4600001	4710000
CP000099_47	4700001	4810000
CP000099_48	4800001	4837408

Continuation (3 of 49) of CP000099 from base 200001 (CP000099 Methanosarcina barkeri str

Query Match 10.1%; Score 102.4; DB 1; Length 110000;
Best Local Similarity 56.1%; Pred. No. 3.2e-08;
Matches 193; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

```
Qy        646 GTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTA 705
              || |  ||    ||||  || || ||    ||  |  ||  | |||| || |    |
Db      82781 GTCATCTTTACAGGGTATGTCTCAGAGTCCGAACTGGCAGAGATATACAATGCTGTGGAT 82722

Qy        706 GGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCC 765
                || || || |||||||||||||||||| || ||  | |||    | || |  |||||
Db      82721 CTGTTTGTATTTCCTTCTTTATATGAAGGGTTCGGGATGCCTCCTCTAGAAGCAATGGCT 82662

Qy        766 TGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCA 825
              ||||| |  |||||  |||||||    ||||| || || || || || |  |||||||||
Db      82661 TGTGGAACACCTGTTATAACTTCAAATACTTCGTCTTTACCTGAAGTTGTAGGAGATGCA 82602

Qy        826 GCGATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATT 885
              || ||  |||| || ||    |  | | |     || || ||| ||| |    |||| |
Db      82601 GCTATAGTTGTAGATCCATATGATGTAAACAAGTTTGCTGAAGAAATGTATGAGTTATTA 82542
```

```
Qy        886 AATGATTCTGAACTTCGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTC 945
              |   ||  ||| ||      |  |  | ||   ||| || ||     || |||| | |||
Db      82541 ATGAATGATGATCTCAAAGAAGAGATGATTAGAAAAGGACTGAAAAGGTCAAAAATGTTT 82482

Qy        946 AATTGGCAAAACGTGGTTAGTGAGATTGAAATGGTACTGACAGA 989
              | ||| | |     ||  |   | |    ||  |||    |  ||
Db      82481 AGCTGGGATACTTCGGCAAAAAAAACATTAAAAGTATATAAGGA 82438
```

RESULT 11
AE013289

```
LOCUS       AE013289               11768 bp    DNA     linear   BCT 07-FEB-2003
DEFINITION  Methanosarcina mazei strain Goe1, section 71 of 379 of the complete
            genome.
ACCESSION   AE013289 AE008384
VERSION     AE013289.1  GI:20905037
KEYWORDS    .
SOURCE      Methanosarcina mazei Go1
  ORGANISM  Methanosarcina mazei Go1
            Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
            Methanosarcinaceae; Methanosarcina.
REFERENCE   1  (bases 1 to 11768)
  AUTHORS   Deppenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,
            Martinez-Arias,R., Henne,A., Wiezer,A., Baeumer,S., Jacobi,C.,
            Brueggemann,H., Lienard,T., Christmann,A., Boemecke,M., Steckel,S.,
            Bhattacharyya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,
            Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.
  TITLE     The genome of Methanosarcina mazei: evidence for lateral gene
            transfer between bacteria and archaea
  JOURNAL   J. Mol. Microbiol. Biotechnol. 4 (4), 453-461 (2002)
   PUBMED   12125824
REFERENCE   2  (bases 1 to 11768)
  AUTHORS   Deppenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,
            Martinez-Arias,R., Henne,A., Wiezer,A., Baeumer,S., Jacobi,C.,
            Brueggemann,H., Lienard,T., Christmann,A., Boemecke,M., Steckel,S.,
            Bhattacharyya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,
            Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute
            for Microbiology and Genetics, Grisebachstrasse 8, Goettingen
            37077, Germany
FEATURES             Location/Qualifiers
     source          1. .11768
                     /organism="Methanosarcina mazei Go1"
                     /mol_type="genomic DNA"
                     /strain="Goe1"
                     /db_xref="taxon:192952"
     gene            complement(90. .599)
                     /gene="MM0645"
     CDS             complement(90. .599)
                     /gene="MM0645"
                     /EC_number="4.2.1.33"
                     /codon_start=1
                     /transl_table=11
                     /product="3-isopropylmalate dehydratase"
                     /protein_id="AAM30341.1"
                     /db_xref="GI:20905038"
                     /translation="MMENPIKGRVWKFGNDIDTDVIIPGKYLRTKDMQVFAAHAMEGI
                     DPGFSKKAKPGDIIVAGDNFGCGSSREQAPLALKHAGIACIVAKSFARIFFRNAINIG
                     LPLMEADIECEEGDQIEVDLLKGEVKVSGKGVFRGNKLPDFLLDMLTDGGLVAHRKKV
                     RDQEKEESA"
     gene            complement(761. .3283)
                     /gene="MM0646"
     CDS             complement(761. .3283)
                     /gene="MM0646"
                     /codon_start=1
                     /transl_table=11
                     /product="Oligosaccharyl transferase"
                     /protein_id="AAM30342.1"
                     /db_xref="GI:20905039"
                     /translation="MNSENHPATTFKSRIKSSLPYTLAVAIIGFVSLWVRMRPYDHVF
                     LSNGFVKFTSNDPWYHIRTLNVLLENYPQRMFFNPMTNYPYGSYIHFGPLYDQMMAIT
                     SLILGLGSPSQDLINTVGAYFPAVLGALTVIPVYYIGKYLGGRKTGILAAVFIAFAPG
```

```
                     QFLSRSLIGFTDHHVAESLFSTFFMMFFMLAIITAKKKNLRFEDLFNKNFIVVKEPLI
                     YSIIAGVMYSAYQLSWPGASLFLFIALVYAVVQYILDNFNGESSDYLGFTGIITFLVS
                     AILILPFIHPELGFSMYYYTWFHVATAIGTMAGFAALSLIQREFKNRNLKAYYYPLAI
                     FLLGFLGLLAIRFASPSVYSLIISAPNTVFGVLTGGAATIGEVSSMFYYGGTFTLSRA
                     FGNFTVSGFFASIIGLIILLVSVIRKAKPEEVLVLVWSILMLFAIYGQNRFAYYYSIN
                     VSILSAYIGGLLLEKVKWNELDEKFKSSVKSPADIPGFLKSFRAKQVLAVLAIAVFLI
                     YPVYGAAMVQSTGSNDPDWAWIEACLWLKSSTPDPGMDYNAIYEAPEDGKLFDYPESA
                     YGVMSWWDYGHYIETLGHRMPNANPFQAGIGGRRGSINETNVPGAAPFLTAQSEEEAT
                     EVLESIHPDPEKSGARYIMSDERMAVDIFMAMPEWTLDTEGYMQPYWTGDGYQYLPSK
                     RYFDSMESRLHFLDGNGLKQYRLVYETWAYQTQEAGYKQVYNFLYGSSIPEVDSGYVK
                     IFEYVKGAKITGTVSPNETVNINTTILTGQGRTFEYSQSTSSDSEGRYEFIVPYSTEG
                     PIPGETQFDTAPTGAYVVSYGDTTTEVRVSEEAVLNGEEIKV"
gene                 complement(3616. .6162)
                     /gene="MM0647"
CDS                  complement(3616. .6162)
                     /gene="MM0647"
                     /codon_start=1
                     /transl_table=11
                     /product="Oligosaccharyl transferase"
                     /protein_id="AAM30343.1"
                     /db_xref="GI:20905040"
                     /translation="MRLLKGINMNSEDYPASVFKNKIKSSLPYTLAVAIIGFVSLWIR
                     TRPSEYVFLSNGFVRFGGNDPWYHMRTLNVLLENYPQRMFFNPMTNYPYGSYIHFGPL
                     FDQMMAIISLILGLGSPSQDLINTVGAYFPAVLGALTVIPVYYIGKYLGGRKTGILAA
                     VLIAFAPGQFLSRSIIGFTDHHVAESLFSTFFMMFFMLAIITAKKNNLRFEDLFNKNF
                     NVVKEPLIYSVIAGVMYSAYQLSWPGASLFLFIALVYAVVQYILDNFNGESSDYLGFT
                     GIITFLVSAILILPFVHPDMGFSLYYYSWFHVATATGIVVCFGILSFIEREFKNRNLK
                     AYYYPLAIFGLGIFGLLAIRIASPPIYSLIINAPHTVFGVQTGGPSTIAEVSSIFYDG
                     GVFTLSRVFGNFTASGFFASLLGMLVLIANAVRKPKPEKVLVLVWSVLILFTIYGQNR
                     FAYYYSINVSILSAYIGGLLLEKVKWNELDEKFKSTVKSPADIPGFLKFLRVEQVLTV
                     LAIVVVLIYPVYGSAMELTKGTGGPDGPWIETCLWLKSYTPDPGMDYNGIYEAPEDGK
                     LFDYPDSAYGIMSWWDYGHWIETIGQRMPNSNPFQAGIGGRRGSMEEENQPGSSTFFT
                     AQSEEEATEVLEAIHPDPEKEGARYIISDIEMATGKFYAMTAWTLDTEGYYQPYWTGS
                     DYQYLPSTRYFDSMVSRLHLLDGNGLKHYRLVHETWAYQTQEAGYKQVYNLLYGSSVP
                     EVDSGYVKIFEYVMGAKITGTASPNETVNINTTILTGQGRTFEYSQSTSSDSEGRYEF
                     TVPYPTEGPIPGETQFDTAPAGAYVVSYGDITKEVRVNEEAVLNGQEIKI"
gene                 complement(6200. .7354)
                     /gene="MM0648"
CDS                  complement(6200. .7354)
                     /gene="MM0648"
                     /EC_number="2.4.1.-"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAM30344.1"
                     /db_xref="GI:20905041"
                     /translation="MASSRNICIVGPSKRFLSGISYYTIRLANAMHSEKDVSVICFRQ
                     LLPTFLFPGKSHVGKDISNLNFSSGIPVFDGMDYNNPLTWLRAYNFIKEQKPDVIILQ
                     WWTSSVAHMQLLLKIFAGMSNKPKIIIEFHEVVDPFEESILPIRLYSKITGKLLRKNL
                     DAYITHSESDKQLVAERYTIAPEKIHVIPHGLYDQYGKLLDTKEAKKNLLINEEFVIL
                     SFGLIRKYKGTPYLIRAFEQLPLEILEKSRLLIVGEIWEDRKELIDQIRASQVRDKIT
                     LVDEYVPDGKVSLYFSASDVVVLPYLRASQSGIAHIAMSFGKPVVVSEVGGLKESMAG
                     YAGTFFVSPGDVDSLRKAILDQFEKKIHYEVPDQKWDRIINRYIELLESI"
gene                 7434. .8558
                     /gene="MM0649"
CDS                  7434. .8558
                     /gene="MM0649"
                     /codon_start=1
                     /transl_table=11
                     /product="glucosyltransferase"
                     /protein_id="AAM30345.1"
                     /db_xref="GI:20905042"
                     /translation="MKIAQICPRYSPDIGGVETHVKEISERLVKAGHDVEVITTDPSG
                     KLEKRDTINGVKVIRFRSFAPGNAYYFAPQIYFYLKKHNFDVIHAHSYHALPALFAAL
                     GKRERKLVFTPHYHRSGHTAFRNLLHKPYRLFGKVIFSRSDSVICVSEYEKRLVEADF
                     RVEGKTVKIPNGINLKEFEHLRNSQKGKGVEKKKGKEKTLLYVGRLEEYKGVQYIIQS
                     LPELKDFRLQVVGNGPYEEELRNMAKNLGVDERVEWLKNLSREELLECYADADIFLML
                     SSHEAYGITVAEALAAGIPCVVAKGSALEEFVDGENCAGIENPVTVEKVIDSLQKIEI
                     FKREPVSGILKGIMDWNEVSARIEKEYIRQSNKTNHKLFN"
gene                 8606. .9667
                     /gene="MM0650"
CDS                  8606. .9667
                     /gene="MM0650"
                     /EC_number="2.4.1.-"
                     /codon_start=1
                     /transl_table=11
                     /product="mannosyltransferase"
                     /protein_id="AAM30346.1"
                     /db_xref="GI:20905043"
                     /translation="MEIDYINGLKTDEIFGMSKYQSEINSRLGNIKLNRIEYPDISKT
                     PGVNKIVEYFVYPFIVKKEVTKNNVKHVTRQDLAFLLELIDLKKTIVTCHDIIPIAYY
                     HTQNPIWKLNAKGLRKAEKIITVSEFSKNDIIKHIRYPEDKIEIIPPAVNHNLYYQNR
                     NKSILKKYGIKEDEKVILYVGAEEPRKNIQVLINSFDKLKSKISQIKLLKVGTPNYLG
                     VREKLLKQIESLNLQKDVIFAGYVPESELAKIYNAVDLFVFPSLYEGFGIPPLEAMAC
                     GTPVITSNSSSLPEVVGDAAIIIDPYDANKFAEKMYEVLINDGLKEEIVRKGLRRSKM
                     FSWEDSARKTLKVYKEFRP"
gene                 complement(9664. .10692)
                     /gene="MM0651"
CDS                  complement(9664. .10692)
                     /gene="MM0651"
                     /EC_number="2.4.-.-"
                     /note="involved in cell wall biogenesis"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAM30347.1"
                     /db_xref="GI:20905044"
                     /translation="MNPKVAIIILNWNGWKDTIECLESVFQIAYPLYEVVIVDNGSKD
                     GSLQKINDYATGKIKTESKFFDYSLENKPIYVKEYSKNELETSVSRSETIKNPASREN
                     LILIKNYDNYGFAEGNNIGIRFALNNLNPDYILLLNNDTVVDPCFLNELINTAENDSL
                     IGVLGPTVYEYKSPEEIQSAGARIYWNKGEVINLTSHESKYPGEPEDVDSVIGCALLA
                     KSELFHKIGYLSKDYFAYMEETEWCVRAHKESYRIVHVPKGKIWHKGGATSNKITGFT
                     LFHHTRNKFWFMKKHSSKKQYFSFLIYFFGFRAWMIIGGILYRQKRKEILPSFISFLK
                     GIRDGILT"
gene                 complement(10689. .11612)
                     /gene="MM0652"
CDS                  complement(10689. .11612)
                     /gene="MM0652"
                     /EC_number="2.4.-.-"
                     /note="involved in cell wall biogenesis"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAM30348.1"
                     /db_xref="GI:20905045"
                     /translation="MLDQKPKVSVIIPTYNRAHLIPRAIKSVLNQTYLDIEIIIVDDS
                     STDNTEEIVKDFKDRRLKYIRHNINKGASAARNTGIRESRGKYIAFQDSDDEWFSNKL
                     EKQIEAFADAPPEVGIVYSGFYRIEANRKLYLPSDQLSLKEGNIHNELLKGNFVGTPT
                     VLMKKECFRNQRYFNESLPALEDWELWIELSKYYTFRYIKKPLLCSYSTPNSINLNEN
                     NMLKAHEIILLTHLDDFSKNKKNLSEHYYDIGTGLCSIGNFNNGKNYLIKSLKAYCLN
                     FKTISVLFILLFGHNIYYKFRDLHRKIGHRT"
ORIGIN

  Query Match             9.9%;  Score 100.2;  DB 1;  Length 11768;
  Best Local Similarity   54.1%;  Pred. No. 1.6e-07;
  Matches  204;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy        613 AAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAA 672
              |||| ||||       |    ||||     |   ||| |  ||    ||||  ||| | ||
Db       9278 AAACAAATAGAGTCCCTAAATTTACAGAAGGATGTTATCTTTGCAGGGTATGTGCCAGAG 9337

Qy        673 AAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTTATATGAA 732
              |  || ||| ||    |||| || |         ||| || || || ||||| |||||
Db       9338 AGTGAATTAGCAAAAATATACAATGCTGTAGACTTATTTGTATTTCCATCTTTGTATGAG 9397

Qy        733 GGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTAACTTCTCTA 792
              ||||||||| ||||     | || |  ||||| ||||| |  || ||| |||| ||
Db       9398 GGTTTTGGAATACCCCCTCTCGAAGCTATGGCTTGTGGGACCCCGGTAATAACATCAAAC 9457

Qy        793 ACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGACCCTCTTTCGGAA 852
              |  || ||| |||| || || |  |||||||| || ||  || | |||||       | |
Db       9458 AGCTCTTCACTGCCTGAAGTTGTGGGAGATGCTGCTATAATTATAGACCCATACGACGCA 9517

Qy        853 GATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTTCGTAAGCATTTA 912
              ||    || || ||  ||| |       | | |   ||  || | |     | | | |
Db       9518 AATAAATTTGCTGAAAAAATGTATGAAGTGCTGATAAATGATGGATTGAAAGAAGAAATT 9577
```

```
Qy        913 ATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTGGTTAGTGAGATT 972
              ||  ||||||  |      |  |||| | ||| |  ||| || |   ||  ||  | ||
Db       9578 GTCAGAAAGGGATTGAGAAGATCAAAAATGTTTAGCTGGGAAGATTCGGCAAGAAAAACG 9637

Qy        973 GAAATGGTACTGACAGA 989
              ||  |||   | |||
Db       9638 TTAAAAGTATATAAAGA 9654


RESULT 12
AE007788
LOCUS       AE007788               10745 bp    DNA     linear   BCT 01-JUN-2004
DEFINITION  Clostridium acetobutylicum ATCC 824 section 276 of 356 of the
            complete genome.
ACCESSION   AE007788 AE001437
VERSION     AE007788.1  GI:15025952
KEYWORDS    .
SOURCE      Clostridium acetobutylicum ATCC 824
  ORGANISM  Clostridium acetobutylicum ATCC 824
            Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 10745)
  AUTHORS   Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
            Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
            Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
            Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
  TITLE     Genome sequence and comparative analysis of the solvent-producing
            bacterium Clostridium acetobutylicum
  JOURNAL   J. Bacteriol. 183 (16), 4823-4838 (2001)
   PUBMED   11466286
REFERENCE   2  (bases 1 to 10745)
  AUTHORS   Childress,D., Zeng,Q. and Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2001) GTC Sequencing Center Production,
            Finishing, and Bioinformatics Teams, Genome Therapeutics Corp., 100
            Beaver Street, Waltham, MA 02453-8443, USA
FEATURES             Location/Qualifiers
     source          1. .10745
                     /organism="Clostridium acetobutylicum ATCC 824"
                     /mol_type="genomic DNA"
                     /strain="ATCC 824"
                     /db_xref="ATCC:824"
                     /db_xref="taxon:272562"
     gene            complement(303. .539)
                     /gene="CAC2904"
     CDS             complement(303. .539)
                     /gene="CAC2904"
                     /codon_start=1
                     /transl_table=11
                     /product="Uncharacterized protein, VEG B.subtilis
                     ortholog"
                     /protein_id="AAK80846.1"
                     /db_xref="GI:15025953"
                     /translation="MIRQNGLASIKASIENHVGEKVTLKSNGGRKKIVVNNGVIEKTY
                     PSIFVVRLESDTQRTVTYSYSDVLTKSVQIVFGA"
     gene            complement(1050. .1952)
                     /gene="CAC2905"
     CDS             complement(1050. .1952)
                     /gene="CAC2905"
                     /codon_start=1
                     /transl_table=11
                     /product="Uncharacterized protein, YABG B.subtilis
                     ortholog"
                     /protein_id="AAK80847.1"
                     /db_xref="GI:15025954"
                     /translation="MVIGDFVVRKSYGEDITFKIIDIINDKGVNYCVLKGVNLRIIAD
                     SPEDDLRLVQASNEGKVDSVFNETINKSIRDIITERSNSIKIGKNILQKTYKKKKMRT
                     SNSSEGINNMTFARPGKVLHIDGDSEYLEVCLKVYKQLSIDAVGKVILEKEQPEKVVD
                     IVKVVKPDIVVVTGHDGMVKNTRNYMDINNYRNSSYFIKSVSNLRDYEPSYDDLVIFA
                     GACQSCYEGILDAGANYASSPGRVLIHCLDPVFVCEKVAYTNVGDIVNIKEVVKGTVT
                     GVKGIGGLQTRGKCREGYPKSSYI"
     gene            complement(2019. .3020)
                     /gene="CAC2906"
     CDS             complement(2019. .3020)
                     /gene="CAC2906"
                     /codon_start=1
                     /transl_table=11
                     /product="Spore coat protein cotS related"
                     /protein_id="AAK80848.1"
                     /db_xref="GI:15025955"
                     /translation="MMREFEIERQFNIKIEYIKPNKGVYFVKTNKGNKCLKKINYGIQ
                     KLLFVYAAKEHLINNGFKYVDKYCLNTEGNPYALVNEDIYTLSEWIEGRECSFYDTED
                     LVKASKTLAKMHECSKGYEPPENSKLKSDLGRWPHLIEKRVKALDKMRDMARKKGNKK
                     DFDMNYIKNVDFFKEQGIKSMNVLKNSKYESLCAVAEEEKGFCHHDYTYHNIIINNED
                     NVYVIDFDYCKRELRAYDISNFMIKVLKRSNWSMKTAELILDSYSSISKLDDDEYGVI
                     YAFLIFPQRFWRLANRYYYNEVNWGQNTFNNKMEELIEERERYIKFIEDFREAYNLQ"
     gene            3812. .4936
                     /gene="CAC2907"
     CDS             3812. .4936
                     /gene="CAC2907"
                     /codon_start=1
                     /transl_table=11
                     /product="Glycosyltransferase"
                     /protein_id="AAK80849.1"
                     /db_xref="GI:15025956"
                     /translation="MKIGIDGRAAKWYRGTGIGTYTYQLISHINKLDEANDYTIFLPD
                     SSNLNFKLKNNFSLKSINCTAKNNFWDQVNIPNIIKKDDTELYHVPQNGIGLPYNKNC
                     RFTITLHDIIPYKMPETVGKKYLKIFEEEIPKIIPLCDGIITVSDFSKKDISEFFNYP
                     ESKIYVTYLSAEDIYRPIEKCISKDIIKNLYGIDKNYILYVGGFSPRKNISGIIHAFS
                     KLKKTTKTNLKLIIAGKKGLSYPNYVDIAKKLNVSSDVIFPGFIPLEYMPFLYNAAEF
                     LIYPSFYEGFGLPPLEAMACKVPVITSNITSLYEIFNGCALIVNPYDVEKIKDAMEIM
                     YFDSSVRKIYSEKGFEFSKKLTWNSTAISTIKAYEKISKN"
     gene            complement(4976. .6010)
                     /gene="CAC2908"
     CDS             complement(4976. .6010)
                     /gene="CAC2908"
                     /codon_start=1
                     /transl_table=11
                     /product="Spore coat protein cotS related"
                     /protein_id="AAK80850.1"
                     /db_xref="GI:15025957"
                     /translation="MLDLKSYREKELASYDLDVELFNEFNFSVYNVMPLRKAFLISTD
                     KGEKILKKIDYTIEEFKFVLAAVNYIKNENEFNRIIDFNKTQSGEYYCIKNGDLYCVM
                     DMIEGKECEFSNPVDLSISSFGLGQLHSASEGFRYSKIASKNICGKMIKGFKRKKEEV
                     NFLKKMVNLYESLNEFDEIFIKNVDYYLEQIDKSIDIMNKSQYYKLCSEEDKIVLCHH
                     DLAHHNIIIKDEEAYFIDFDYAIIDLKVHDLSNIISKAAKVFNYSIDNSKIIISNYCK
                     SNTLSKNEMEVLFGMLSFPEDFYEIVKSYYTKRKDWSNEEFINKMKRKMANEENRVEF
                     LKNFKEDLKI"
     gene            complement(6022. .6777)
                     /gene="CAC2909"
     CDS             complement(6022. .6777)
                     /gene="CAC2909"
                     /codon_start=1
                     /transl_table=11
                     /product="Spore coat protein cotS related (diverged)"
                     /protein_id="AAK80851.1"
                     /db_xref="GI:15025958"
                     /translation="MGSLDDISFIDYIEENGIRLSNIVNHIDKKDIDEKKAIEQIQII
                     ADFHRIARGYNGLDVKTIKNKIGKLVGQYKIHAKKLKKYADMKKKSSNLTMFEEYFLN
                     TCESVIERCEKCLGHINDDEYMEIIKRSMRNNDICLGNCYFENLWKNGNIFLADYSDI
                     AFNMVEIDGVKFIGKLKSNSLDLNFELLIDKYIEAENIKSESKNLMKALISYPSEYIK
                     CVERYERNSDESLEDRFVKRLAKSIRKDGESII"
     gene            complement(6792. .7799)
                     /gene="CAC2910"
     CDS             complement(6792. .7799)
                     /gene="CAC2910"
                     /codon_start=1
                     /transl_table=11
                     /product="Spore coat protein cotS related"
                     /protein_id="AAK80852.1"
                     /db_xref="GI:15025959"
                     /translation="MNSVEINEFLKNNYNINAEASQKIKNVYRVNSSSKNYCLKCIHY
                     DYGHVLFIVKAMEHLKENGFDKIPEIIPTSSGEKFIKFRSCYVYLTEWLDARVCNFDN
                     PIDLYTAVKALSSLHLKSRSFEVTDNMNPRVGWFNWRENFKRRKDEILKFKKIICDKE
                     KKTEFDNLYMKLMEEELSKADKSIQDIENSEYVNKMNEEFRLKGFCHHDYAHHNILID
```

```
                     KEGKAHIIDFDYCMLDTHLHDLSSILIRKMKNGLWDMKCCLFILNTYNEDYKIENNDI
                     PIMASFMEFPQGYWQVGLQYYVEKQPWTEEFFMSRLNKIYKDKDERQEFLEDFMKFKY
                     L"
     gene            7971. .9095
                     /gene="CAC2911"
     CDS             7971. .9095
                     /gene="CAC2911"
                     /codon_start=1
                     /transl_table=11
                     /product="Glycosyltransferase"
                     /protein_id="AAK80853.1"
                     /db_xref="GI:15025960"
                     /translation="MNIAIDARGINLYRGTGIGTYTENLVRELLNIDEENSYTLYWYG
                     NNYKDFIKSNTKINIVSKGCHSFFEKCYFPENIIKNSIEIYHVPQNGIGLDQNVSCIK
                     VSTVHDLIPYIMPETVGKGYLLKFLKNMPFIIENSDAILTVSEYSKKDILKYFPINED
                     KIYVTPLAANYNYKPLSREHCKKFIEKNYSISDPFVLYLGGFSKRKNVRNLILAFADA
                     SKKLSKKHNLVIIGLCRDELESLKDLCRHLNISNEVIFTGYIGERFLPVFYNACELFV
                     YPSLYEGFGLPVLEAMSCKTPVVTSNISSIPEIAGDSAVLINPLDTSELRDAILNILE
                     DSKLKQKLSIEGFNRSRKFSWKNTSKKTLEAYSNIYNTFA"
     gene            complement(9316. .10614)
                     /gene="CAC2912"
                     /note="araN"
     CDS             complement(9316. .10614)
                     /gene="CAC2912"
                     /codon_start=1
                     /transl_table=11
                     /product="Sugar-binding periplasmic protein"
                     /protein_id="AAK80854.1"
                     /db_xref="GI:15025961"
                     /translation="MKRFLKKIFVAVMCSSMVFYVSACTNTLNNTSPVKHDKKVNQKK
                     SSITIWINNKDQNIIKDQINAFNKKYDYITVNTVFMSEDDIIKNYENNTKDNLKLPDV
                     VEIGDLSASKFVNKFSGKLVNLSSDSDIKNDMFLSNKIHNLTKNSRVYGYPWYTQPVF
                     MFYRSDIMHSLNINTDDIKTWKDYNDLGNKVKQNGKNLIYLSDLSDIYDVQNSELGVN
                     FLDNNDKINIEDNKFKEAASFVFDIIKGNTNISIQPGKDITKAFASGEIVSMMSTPEE
                     ILYLEKNYPNLKGKVSLERLPAFEQGGNNTAYKDGMNLLVTDNEKNKTYSEQFAKFVT
                     SNEKAAYNQFSIYALCSSNSSAYSYDKMHEKNDYMGGCDVYEMYMNAAKMFSDVKYYK
                     DYSDIKNYIVKEIGSEAAQNKGIDEIITNMQKTLADRYNK"
ORIGIN

  Query Match             9.9%;  Score 100;  DB 1;  Length 10745;
  Best Local Similarity   53.6%;  Pred. No. 1.8e-07;
  Matches  208;  Conservative    0;  Mismatches  180;  Indels    0;  Gaps    0;

Qy        570 AATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACA 629
              | ||| | | | |   | |  ||     | |   |||| || |  ||   | ||| |
Db       8654 AGTAATAATAGGTCTGTGTAGAGATGAACTTGAGAGTTTAAAAGACTTATGTAGACATCT 8713

Qy        630 TGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTT 689
              | ||  | |    || |  ||    || | | |    |||| |  ||||||    ||
Db       8714 AAATATATCTAATGAAGTTATCTTTACAGGATACATTGGAGAAAGATTTTTACCTGTATT 8773

Qy        690 ATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGT 749
              ||||| |  |     | ||| || |  |||||| | |||||||||||||  | || ||
Db       8774 CTATAATGCTTGTGAATTATTTGTGTATCCTTCTCTGTATGAAGGTTTTGGCCTGCCGGT 8833

Qy        750 AGTGGAGGGCATGGCCTGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGA 809
              | | || |  |||  |||     |  |||||  |||||||     |  || ||| | |||||
Db       8834 ACTTGAAGCAATGAGCTGCAAAACACCTGTTGTAACTTCAAATATATCTTCAATACCAGA 8893

Qy        810 GGTGGCTGGAGATGCAGCGATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAAGG 869
              | |||||||||  | ||  | ||| |  | ||||||          |   |||   | |
Db       8894 AATAGCTGGAGATTCTGCTGTACTTATAAATCCTCTTGATACTTCAGAACTTAGAGATGC 8953

Qy        870 AATTTCGAGGTTAATTAATGATTCTGAACTTCGTAAGCATTTAATCCAAAAGGGGCTTTT 929
              ||| |  |    |  || | |||  | || |      |  | |||| |  | | ||  ||
Db       8954 AATATTAAATATTCTTGAAGATAGTAAATTAAAACAAAAGTTAAGCATAGAAGGTTTTAA 9013

Qy        930 GCGGGCAAAGAGGTTCAATTGGCAAAAC 957
              ||  | |  |  ||    ||| |||||
Db       9014 TCGCTCCAGAAAATTTTCATGGAAAAAC 9041


RESULT 13
AB046360
LOCUS       AB046360               15890 bp    DNA     linear   BCT 20-JUL-2000
DEFINITION  Actinobacillus actinomycetemcomitans serotype a-specific
            polysaccharide antigen gene cluster (manC, orf3, orf4, orf5, orf6,
            orf7, orf8, gmd, orf10, orf11, orf12, orf13), complete cds.
ACCESSION   AB046360
VERSION     AB046360.1  GI:9309318
KEYWORDS    .
SOURCE      Actinobacillus actinomycetemcomitans
  ORGANISM  Actinobacillus actinomycetemcomitans
            Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
            Pasteurellaceae; Actinobacillus.
REFERENCE   1  (sites)
  AUTHORS   Suzuki,N., Nakano,Y., Yoshida,Y., Nako,H., Yamashita,Y. and Koga,T.
  TITLE     Genetic analysis of the gene cluster for the synthesis of serotype
            a-specific polysaccharide antigen in A. actinomycetemcomitans
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 15890)
  AUTHORS   Suzuki,N., Nakano,Y., Yoshida,Y., Nakao,H., Yamashita,Y. and
            Koga,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2000) Nao Suzuki, Kyushu University Faculty of
            Dentistry, Depertment of Preventive Dentistry; 3-1-1 Maidashi,
            Higashi-ku, Fukuoka-shi, Fukuoka 812-8582, Japan
            (E-mail:nao@dent.kyushu-u.ac.jp, Tel:81-92-642-6423,
            Fax:81-92-642-6354)
FEATURES             Location/Qualifiers
     source          1. .15890
                     /organism="Actinobacillus actinomycetemcomitans"
                     /mol_type="genomic DNA"
                     /strain="SUNYaB 75"
                     /serotype="a"
                     /db_xref="taxon:714"
     gene            755. .2158
                     /gene="manC"
     CDS             755. .2158
                     /gene="manC"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative GOP-mannose phosphorylase"
                     /protein_id="BAB03201.1"
                     /db_xref="GI:9309319"
                     /translation="MFISVIMAGGSGSRLWPLSRSMFPKQFLALDKKNNLTMLQATID
                     RINQMSTSEPIIITNEQHRFIVAEQLREYGVKSRIILEPVGKNTAPAIALAAFEAIKH
                     GEDPILLVLAADHVVLNKEAFQTSVRNALVQAQTGYLVTFGIVPTAPETGYGYIQRGE
                     KITEEAYKVERFVEKPNFETAEKYVASGNCYWNSGCFMFKASCYLHELRKFALEIYEA
                     CEKATSNRLKNNDFIRVDVDEFLKCPDDSVDYAVIEKTDKAIVVPMDAGWSDVGSWSA
                     LWEVSDKDENQNSYHGDVIMKNTSGCYIYAPNKLVAAVGLENIVIVDTKDAVLVADRS
                     KVQEVKQIVEQLKKEDRREYKEHREIYRPWGRSDAIAQGERYKVNRITVNPGESEETQ
                     VHYHRAEHWIVVSGMAKVTKGEQESLITENQSIYIPVGVPHQIANLGKIPLELIEVQS
                     GGYLDDDDVSVLRKYYK"
     CDS             2182. .2973
                     /note="orf3"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative ABC-transporter"
                     /protein_id="BAB03202.1"
                     /db_xref="GI:9309320"
                     /translation="MLNKFISNIKDFYKFNELLKQLVIRDIKLKYRRSYLGYLWSILN
                     PLMLMLVLVVVFSNLFRFDIPNFPTYLITGQIIFGFMNEATNVAVGSIINNSSLINKI
                     YVPKYIFTLSKVCSSLVNLIFSFGALIIVLIMTQTPVSWNILFVPIVILQVFLFSLGL
                     SLFISAFTVFFRDIQYLWGVFVSMWMYFTPLFYPISIISEEYQGVYKALNPMYWYIEQ
                     FREIILFENLPQTESIIYGVLYAILVLFLGILYFNKRQNEFILYI"
     CDS             2970. .3722
                     /note="orf4"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative ABC-transporter"
                     /protein_id="BAB03203.1"
                     /db_xref="GI:9309321"
```

```
                     /translation="MKEFEKNNIIEVCNATVRFNRTSESYNGLKEYVIKMLKKELMFQ
                     EFLALEDISFSVKKGEAWGILGINGSGKSTLLKLICGIVRPYKGTVKTEGKIAPLIEL
                     GAGFDHELTARENIFLNAALLGYNKSFALQCFDEIIEFSELYDFVDVPIKNFSSGMVA
                     RLGFAIATMVKPDILIIDEVLAVGDAAFQKKCKHRMEKMLSDGTTLLFVSHSTQQVAE
                     LCRKAIWLDKGKIKAIGEVNEVINKYINLLTL"
     CDS             3728. .5728
                     /note="orf5"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative glycosyltransferase"
                     /protein_id="BAB03204.1"
                     /db_xref="GI:9309322"
                     /translation="MLRYFRLFFIWEGIMRIWFDISNISYWQGGIVGVIRAELELAYN
                     FSKISDNILFCRVNSSGDAIEAVTREELSWLLNASTAVDGFIQYQKVYKESKKLPLSH
                     FPKTVDNFYSTNSKLQHISLAIKFFSSLLPMYLGNKVYKFSIALENTVNRYRSKYSKI
                     LKISEKENKNKTDNPSYPFYSGDVIFSAGWKDSNKENLYKRIKMHLNGNLYLSYLIYD
                     TILINKNTKHLYSKETESSFYRYFSWISENCDLIFYGGKTAMLDSQEIQKQKRLSIPK
                     GIPIRFGDLPNPNLKSLTDSEITQLFDKLNIVKNKYLLCVGSIEPRKNHSILYKAYRE
                     IIDNPDKYNFSLENLPKLVFVGKLYGSEELVYNFRNDPKLKSLVVFISPTDEELDSLY
                     KYCIFTLMPTLYEGWNLTMPESLSYNKFCISSNVAPMIEIGKDLVEYADPYNPIEWAK
                     LIAKYCSNNSLLFEREEFIRQYWKNYSWYECAKFIFESLNNEDFPLQNMRTTLWYDLT
                     LVNNLYRKLDGIPRVQLSMAWELYHKDNLNIRFFEFSSNQNEGTFVELPVNTVPWLSK
                     PREEYVDFYYTHQNYKQHIYSLQHIREHNESIESGIINIKNLEPSAKRRLRISLLHVL
                     GVLPISVLKILYSVYAKLIKKEEEKSPQIYLKIHLQNYLLTL"
     CDS             5797. .6705
                     /note="orf6"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative glycosyltransferase"
                     /protein_id="BAB03205.1"
                     /db_xref="GI:9309323"
                     /translation="MGLDWGSRYLQGIDVLKKTNRISSIYFIYDMIPLLEPNFYEDYI
                     HGLYEEFFYWVCKTSDQILFGGKTALNDGKNIQKKLKIDSNGSMSFLRLGSEFSLSES
                     ASLDKNNNELKKLGIDRPFILSVGTFQIRKNHEVVYRALLQWLKETPEEEHKNLPIFV
                     FAGRQGWLVKQLVKEMESDERVQRHLILMHPSDDELDILYKNCLFTVLPSLYEGDSLA
                     MAESLACGKLCIASDVPPLRETGEGFADFVEPRNPQAWKEKCEHYFKNPKLLRQREKL
                     IRNSRAFISWDECADNLLNKILNSNR"
     CDS             6718. .7602
                     /note="orf7"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="unknown"
                     /protein_id="BAB03206.1"
                     /db_xref="GI:9309324"
                     /translation="MKILVTGGSGFIGKNLIYLLREKREFEVFGATVEETMDLTNPCS
                     VQSVLEKTKPDFIVHLAALTFVPNNNPITFYLVNTIGTENLLRSIVDLNVAKLGVLCF
                     STAGIYGIQETKLLSESLTPKPVNHYSMSKHCMEHIVNKYRCFRGITVVRPFNVLGLG
                     QNINFLVPKMVSAFVKKDKTIELGNLDSVRDFISVNDCCDIIYRLISKLIENETINIC
                     TGIGYSVYQIFQLLCEISMHQMEIKQNELFVRHDDIPQMIGDPSKLLNVLGNDYRFTS
                     VRAILEEMYKNRLLELSI"
     CDS             7604. .8797
                     /note="orf8"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative acetyltransferase"
                     /protein_id="BAB03207.1"
                     /db_xref="GI:9309325"
                     /translation="MIFEIFVLFYSISFIPICVWYTGYILKIVNVKFPEYNSSFNKPL
                     ESLRGIACAAVLMAHINMYFYSWTNSYALKEFSSSFKSSSYLGQSGVVWFFMLTGYLF
                     WNMVQSNSFDMNNFYIKRIRRLVPVMIVLVLVVSIFDWVMGGTPIPNLEQLVKLLKNF
                     SFYFAGIHDTFYPEIVHRINNLWTLRWEWMFYLVLPLISTLCRSWLALTILFFPSIFI
                     LTDVYYTGVKQSDAALFLAFYLGMIFVHFIKFIKVTNFLKLKKIILLSLSPYGFLFVI
                     LLSMTSMFMPASEIRPQNIYFIFINAPVFLWFLLYSVFDNDRMKWILSLPCLNLGRIS
                     YSLYLWHLFITFIVYKLFINIPPFINNFFSLIFLYPVVVILVSIPISILSYKLIELRFM
                     YKHMK"
     gene            8830. .9864
                     /gene="gmd"
     CDS             8830. .9864
                     /gene="gmd"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative GDP-mannose dehydratase"
                     /protein_id="BAB03208.1"
                     /db_xref="GI:9309326"
                     /translation="MKTAIVTGASGQDGAYLSQLLLDKGYKVYATYRRSSSVNLWRID
                     ELNIRNHPNLHLFEFDLTDMSSCISLVTKAQPGEVYNLAAQSFVGVSFSQPVTTAEIT
                     AIGVLNLLEAIRIINPKIKFYQASTSEMFGKVQQIPQTEKTPFYPRSPYGVAKLYGHW
                     ITLNYRESYDIFGCSGILFNHESPLRGREFVTRKITDTVAKIALNKQSCLELGNLDAK
                     RDWGFAKEYVEGMWRMLQEDQPDTYVLATNRTETVRDFVAMAFQAVNIPLEFNGKGEN
                     EIGVNTDTGDVLVRVNKEYYRPAEVDLLIGDYSKAKRILGWEPKTSLEELCKMMIEAD
                     IERNKLGFSF"
     CDS             9953. .11125
                     /note="orf10"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative glycosyltransferase"
                     /protein_id="BAB03209.1"
                     /db_xref="GI:9309327"
                     /translation="MTKLIINLNVLNYPLTGIGYYTLNIVKELLNRNVELVGVKDSNI
                     LNKEHIFSLINQDIKGKKYSTTKKNLKIMVVNLLRKIPKIYYIRCKIKSLLLSTKMNV
                     LAEKGYVYFEPCFVPLGYSGSIITTIHDLSFIQYPEFHPEERVKYLTSELPRTVRLSS
                     IIFVDSDAIAVELKKLYPMYSNKVVTLYPGVDTSFKQYENCNDVLSSFGLKDKSFILS
                     VCTLEPRKNLLRLVEAYKLLPESMKEEYPLVLVGDQGWKNSELFNSAKELFDKKQLIM
                     TKYISDDLVKKIYSSAAIFAYPSLYEGFGLPIVEAMASGTAVITSNYGAMAEVAGDSA
                     VLVNPHSTEELKDAIINLLTDIKFRNYFIDRGLNRAKIFSWNNTVDKLISVVDEIN"
     CDS             11127. .12218
                     /note="orf11"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative glycosyltransferase"
                     /protein_id="BAB03210.1"
                     /db_xref="GI:9309328"
                     /translation="MKVAIIHYWLVNLRGGEKVLEALCELYPQADIFTHVYIQEKFQN
                     SVISKHKITTTFISKLPWAAKHYQLYLPFMPLALEEMDLSGYDLVISSESGPAKGVIV
                     PPGVPHICYCHSPMRYAWDMYGEYRAKMGKIKRFLAALFLHYIRRWDQLTSQQVTKFI
                     ANSQFVAKRIKSYYAQESMVIYPPVQVDAFELAQDSEDYYLILGQLVPYKKTDLAVRA
                     FNQSGKKLVVIGDGVQLELLKKIAQRNVKILGYQPFQVVKEYLMKCKALIFPGVEDFG
                     IVPLEAMACGKPVIAFAQGGALETIKEGITGSFFYNQTENSLNWAVSHFEQNFYVEPY
                     IIRQHAMLFSKENFKEKIHSYINQIITGK"
     CDS             12229. .13647
                     /note="orf12"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="putative glycosyltransferase"
                     /protein_id="BAB03211.1"
                     /db_xref="GI:9309329"
                     /translation="MNNSAYPKINKISLVATDFFVLCFSIFIAYYFSSPDDFSWDVDK
                     FMDSIIRLNLFLFTAFGVIAWIWTVYRHYSYRKPFWDELRDIYVTIFIASFINLALFV
                     FSQNNFSVLTWLYVWGSALILFPVFRILIKKLLRKCGTWDLNCVLIGNTENAVHGYQA
                     ISSESDLGYKVTAFVNTSNRSHTLEEMGLKLITEEQLITTINDYQKVFIALNNNQNDL
```

```
Query Match             9.8%;  Score 98.6;  DB 1;  Length 15890;
Best Local Similarity   62.2%;  Pred. No. 2.7e-07;
Matches  155;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        708 ATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTG 767
              ||| | || |||||| |||||||||| ||||| |||||  |||||||  |  |||||  |
Db      10834 ATTTGCTTACCCTTCATTATATGAAGGATTTGGTTTACCAATAGTGGAAGCAATGGCAAG 10893

Qy        768 TGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGC 827
              |||||   |||| | || ||      |  | |  || |||| || || || ||| ||||
Db      10894 TGGTACAGCTGTAATTACCTCAAATTATGGAGCTATGGCAGAAGTCGCCGGGGATTCAGC 10953

Qy        828 GATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAA 887
                || | ||| | || | |||    || |  | |   | | |||    |  ||| | |
Db      10954 TGTTTTAGTCAATCCCCATTCAACTGAAGAATTAAAAGATGCAATAATTAATTTATTAAC 11013

Qy        888 TGATTCTGAACTTCGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAA 947
```

```
                ||| | || | ||||| |||| ||  | |  ||  |     |||| |||| |||| 
Db      11014 AGATATTAAATTCCGTAACTATTTCATAGATAGAGGATTAAATAGGGCTAAGATTTTCAG 11073

Qy        948 TTGGCAAAA 956
              |||| | ||
Db      11074 TTGGAATAA 11082


RESULT 14
AE009578/c
LOCUS       AE009578               11501 bp    DNA     linear   BCT 20-MAR-2003
DEFINITION  Brucella melitensis 16M chromosome I, section 135 of 195 of the
            complete sequence.
ACCESSION   AE009578 AE008917
VERSION     AE009578.1  GI:17983399
KEYWORDS    .
SOURCE      Brucella melitensis 16M
  ORGANISM  Brucella melitensis 16M
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Brucellaceae; Brucella.
REFERENCE   1  (bases 1 to 11501)
  AUTHORS   DelVecchio,V.G., Kapatral,V., Redkar,R.J., Patra,G., Mujer,C.,
            Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A., Lykidis,A.,
            Reznik,G., Jablonski,L., Larsen,N., D'Souza,M., Bernal,A.,
            Mazur,M., Goltsman,E., Selkov,E., Elzer,P.H., Hagius,S.,
            O'Callaghan,D., Letesson,J.-J., Haselkorn,R., Kyrpides,N. and
            Overbeek,R.
  TITLE     The genome sequence of the facultative intracellular pathogen
            Brucella melitensis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
   PUBMED   11756688
REFERENCE   2  (bases 1 to 11501)
  AUTHORS   DelVecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2001) Institute of Molecular Biology and
            Medicine, University of Scranton, Scranton, PA 18510, USA
REFERENCE   3  (bases 1 to 11501)
  AUTHORS   Elzer,P.H. and Hagius,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag
            Center, 111 Dalrymple Building, Baton Rouge, LA 70803, USA
REFERENCE   4  (bases 1 to 11501)
  AUTHORS   Kapatral,V., Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A.,
            Lykidis,A., Reznik,G., Jablonski,L., Larsen,N., D'Souza,M.,
            Bernal,A., Mazur,M., Goltsman,E., Selkov,E., Haselkorn,R.,
            Kyrpides,N. and Overbeek,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell
            Park Drive, IL 60612, USA
REFERENCE   5  (bases 1 to 11501)
  AUTHORS   Letesson,J.-J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2001) Unite de Recherche en Biologie Moleculaire,
            Laboratoire d'Immunologie et de Microbiologie, Universite of Namur,
            61 rue de Bruxelles, Namur 5000, Belgium
REFERENCE   6  (bases 1 to 11501)
  AUTHORS   O'Callaghan,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue
            Kennedy, Nimes 30900, France
FEATURES             Location/Qualifiers
     source          1. .11501
                     /organism="Brucella melitensis 16M"
                     /mol_type="genomic DNA"
                     /strain="16M"
                     /db_xref="taxon:224914"
                     /chromosome="I"
     gene            113. .667
                     /gene="BMEI1401"
     CDS             113. .667
                     /gene="BMEI1401"
                     /codon_start=1
                     /transl_table=11
                     /product="TRANSPOSASE"
                     /protein_id="AAL52582.1"
                     /db_xref="GI:17983400"
                     /translation="MREQSSELFSCRTVDEVSIHRSGEEGIPHISSLPCHGRKPKWLF
                     RLEYRPVRRRQRDDLVMLAHMRTAWSLGRHRIARLIRENAMQARQKRRFRKTSDSHHA
                     FPVAPNILAQDFAATGPNQKWDADISYIWTKEGWLHLAGVIGLFARRVVGWSVSDRLH
                     RQLALQALEKAITMRRPPKGLIHH"
     gene            complement(757. .1197)
                     /gene="BMEI1402"
     CDS             complement(757. .1197)
                     /gene="BMEI1402"
                     /codon_start=1
                     /transl_table=11
                     /product="TRANSPOSASE"
                     /protein_id="AAL52583.1"
                     /db_xref="GI:17983401"
                     /translation="MNSKLHAVCDGHGRPLILLLSEGQMSDYKGAAKMLHAFPKAKDL
                     LADKGYDADWFRDALAERKITACIPSRANRKDVIPYDPVLYKRRHKIENMFGRLKDWR
                     RIHTRYDRCAHTFFSSICIAAAVTVLQINSDFNVLKNVSTIALS"
     gene            complement(1194. .1562)
                     /gene="BMEI1403"
     CDS             complement(1194. .1562)
                     /gene="BMEI1403"
                     /codon_start=1
                     /transl_table=11
                     /product="TRANSPOSASE"
                     /protein_id="AAL52584.1"
                     /db_xref="GI:17983402"
                     /translation="MSDLLLLSQAQVCRIKPYFPLSHGVPRVDDRLILSGIIFVLRTG
                     LRWRDAPREYGPHTTIYNRFIRWSRLGVFNRILAELAAKRGKPDQLMIDATHLKAHRT
                     AASLLKKGMFPDVLGAPKVA"
     gene            complement(1620. .2738)
                     /gene="BMEI1404"
     CDS             complement(1620. .2738)
                     /gene="BMEI1404"
                     /EC_number="2.4.1.-"
                     /codon_start=1
                     /transl_table=11
                     /product="MANNOSYLTRANSFERASE"
                     /protein_id="AAL52585.1"
                     /db_xref="GI:17983403"
                     /translation="MRIGVDARNLVNNITGISRYVMENCRELYERGHELVLYTPEKPR
                     VQLSSSEGFDFRVSNFQGPAGRLFWSQTVLPLQLRKDRIDVFWGPAHRLPALKGAIPS
                     VVSIHDLVWYYASSTMRLQGWLADRFLMKGALRNADQIVAVSYATSSAISSVFPQYTS
                     KIQMVYPGVSHFNRIGTSNILKEHRINKSYGLFIGTLEPRKNLIRLLEAYANLPECVR
                     NNFLLVVAGGKGWNLGDIGKTINRLGIAEYTRLTGYVTDSDLNDLYRNARTLLMPSLY
                     EGFGLPIIEAQAYGVPVITSNCSSMPEVAGDAAILVDPLNVAEISNALYRIATDNKLW
                     QRKSLLALQNIERFNWHNSIEKLESVFIKARESKLMTY"
     gene            complement(2869. .3108)
                     /gene="BMEI1405"
     CDS             complement(2869. .3108)
                     /gene="BMEI1405"
                     /codon_start=1
                     /transl_table=11
                     /product="TRANSPOSASE"
                     /protein_id="AAL52586.1"
                     /db_xref="GI:17983404"
                     /translation="MQTRPKAAERKAWANIPPKSNRKDSFVFSRWVCRQRSLVERFFN
                     RIKQFRDIATRYDKRPENYLAAVKLVATRIWCQSL"
     gene            complement(3224. .3502)
                     /gene="BMEI1406"
     CDS             complement(3224. .3502)
                     /gene="BMEI1406"
                     /codon_start=1
                     /transl_table=11
                     /product="TRANSPOSASE"
                     /protein_id="AAL52587.1"
                     /db_xref="GI:17983405"
                     /translation="MTTKLHAVVDAIGLPLRIKPTPGHYGDCPQASSLLSGLEGVGHV
                     IADAAYDADHLRAFIASDLKATAQIQGQSNTFQCPNNRLEAVQGTPLD"
     gene            complement(3499. .3867)
                     /gene="BMEI1407"
     CDS             complement(3499. .3867)
```

```
                     /gene="BMEI1407"
                     /codon_start=1
                     /transl_table=11
                     /product="TRANSPOSASE"
                     /protein_id="AAL52588.1"
                     /db_xref="GI:17983406"
                     /translation="MSRRSLTDEQWNRIEAYLPGRVGTPGRSGVDNRLFVDAILWMAA
                     NAAHWRDLPATFGKWTAVHARFRRWSHAGVWERLFHALADTPDFEYVLIDSTISKVHA
                     DAAGAKGGLKLPASVARAGD"
     gene            complement(3937. .4347)
                     /gene="BMEI1408"
     CDS             complement(3937. .4347)
                     /gene="BMEI1408"
                     /codon_start=1
                     /transl_table=11
                     /product="TRANSPOSASE"
                     /protein_id="AAL52589.1"
                     /db_xref="GI:17983407"
                     /translation="MIRTVPYRIHTILTDNGVQFVQREQGTGLTFSYIFGRVALEHGI
                     EHRLTKPYHPWTNGQAERMVRTIKEATVKSFHYSSINEMRRHVRDWLTAYNFAKQLKA
                     LKFKTPYEAIHELWKSKPEAFIVKPHHHMMGPNT"
     gene            complement(4344. .4730)
                     /gene="BMEI1409"
     CDS             complement(4344. .4730)
                     /gene="BMEI1409"
                     /codon_start=1
                     /transl_table=11
                     /product="TRANSPOSASE"
                     /protein_id="AAL52590.1"
                     /db_xref="GI:17983408"
                     /translation="MPMGPKERRSTVLAPVEEAAIVALRVQARLPLDDVYIALKDVIP
                     HLTCSSLHRCLQRHGISRLPKGDREKPKKFKDYEIGYFHIDIAELRYEGGKGFLFVAV
                     DRTSKLVFARIYRKATKLAPPGSSSH"
     gene            complement(4760. .5134)
                     /gene="BMEI1410"
     CDS             complement(4760. .5134)
                     /gene="BMEI1410"
                     /codon_start=1
                     /transl_table=11
                     /product="TRANSPOSASE"
                     /protein_id="AAL52591.1"
                     /db_xref="GI:17983409"
                     /translation="MGRSRGGLTSKIHALVDAEGRPVNLRLTGGQIADCAEADALTNE
                     FGEGDILLADKGYDTNAIRARCLVPACDGVELSPTQRRELRDKFFTAAPRPRTPFEQS
                     YSDRKLRPRSLRGGTASMRKRS"
     gene            complement(5053. .5361)
                     /gene="BMEI1411"
     CDS             complement(5053. .5361)
                     /gene="BMEI1411"
                     /codon_start=1
                     /transl_table=11
                     /product="TRANSPOSASE"
                     /protein_id="AAL52592.1"
                     /db_xref="GI:17983410"
                     /translation="MAISNGLALGGSSRTLWPAAACYNRFVRWRKAGVWDRLFETVPK
                     AYDGDIVMIDSTCVRVHQHAATGKKGMETMVAWDVPAAGLRAKSTRSLMPKVARSIYV
                     "
     gene            complement(5205. .5480)
                     /gene="BMEI1412"
     CDS             complement(5205. .5480)
                     /gene="BMEI1412"
                     /codon_start=1
                     /transl_table=11
                     /product="TRANSPOSASE"

  Query Match             9.6%;  Score 97.4;  DB 1;  Length 11501;
  Best Local Similarity   52.9%;  Pred. No. 4.8e-07;
  Matches  209;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy        604 GATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTC 663
              ||| | |  ||    |||| |  | ||| |   |||   |       ||   |||||
Db       2036 GATATTGGGAAGACTATAAATCGGCTTGGCATTGCTGAGTATACGCGTTTAACTGGGTAT 1977

Qy        664 GTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCT 723
              ||  | || |  ||| ||      |||||||      |        |  |  | || |||
Db       1976 GTTACAGACAGCGATCTAAACGATTTATATAGAAATGCACGGACGCTTCTCATGCCGTCT 1917

Qy        724 TTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTA 783
              | || |||||||||| |||||  || | ||||     ||||  ||  |||||||  |
Db       1916 CTCTACGAAGGTTTTGGCTTACCCATAATTGAGGCACAAGCCTACGGAGTTCCTGTGATT 1857

Qy        784 ACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGACCCT 843
              || ||      ||||||  |||| ||||||||||| ||||||| ||  | ||||| |||
Db       1856 ACGTCAAACTGTTCATCGATGCCTGAGGTGGCTGGGGATGCAGCAATCTTGGTCGATCCT 1797

Qy        844 CTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTTCGT 903
              ||    | ||  |  ||  | || | | |||  ||  |   || ||||  | ||||  |
Db       1796 CTGAACGTAGCCGAAATATCGAATGCACTTTATAGAATCGCTACTGATAATAAACTATGG 1737

Qy        904 AAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTGGTT 963
              |||           |   |   |           || | ||||| ||||| ||     |
Db       1736 CAGCGGAAGTCATTACTTGCTTTACAAAACATTGAGCGTTTCAACTGGCATAATAGTATA 1677

Qy        964 AGTGAGATTGAAATGGTACTGACAGAGGCATGTGA 998
                || | ||||  ||  | |   | ||  | ||
Db       1676 GAAAAGCTGGAAAGTGTTTTCATCAAAGCCAGAGA 1642


RESULT 15
AF047478/c
LOCUS       AF047478               13835 bp    DNA     linear   BCT 05-JUN-2000
DEFINITION  Brucella melitensis strain 16M lipopolysaccharide O side chain
            biosynthesis gene cluster, complete sequence.
ACCESSION   AF047478 AF036614
VERSION     AF047478.1  GI:4071211
KEYWORDS    .
SOURCE      Brucella melitensis
  ORGANISM  Brucella melitensis
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Brucellaceae; Brucella.
REFERENCE   1  (bases 1 to 13835)
  AUTHORS   Godfroid,F., Taminiau,B., Danese,I., Denoel,P., Tibor,A.,
            Weynants,V., Cloeckaert,A., Godfroid,J. and Letesson,J.J.
  TITLE     Identification of the perosamine synthetase gene of Brucella
            melitensis 16M and involvement of lipopolysaccharide O side chain
            in Brucella survival in mice and in macrophages
  JOURNAL   Infect. Immun. 66 (11), 5485-5493 (1998)
   PUBMED   9784561
REFERENCE   2  (bases 1 to 13835)
  AUTHORS   Godfroid,F. and Letesson,J.-J.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1998) Immunologie et Microbiologie, FUNDP, Rue de
            Bruxelles, 61, Namur 5000, Belgium
COMMENT     On Dec 30, 1998 this sequence version replaced gi:3170022.
FEATURES             Location/Qualifiers
     source          1. .13835
                     /organism="Brucella melitensis"
                     /mol_type="genomic DNA"
                     /strain="16M"
                     /db_xref="taxon:29459"
     gene            complement(961. .2091)
                     /gene="wbkA"
     CDS             complement(961. .2091)
                     /gene="wbkA"
                     /note="WbkA"
                     /codon_start=1
                     /transl_table=11
                     /product="mannosyltransferase"
                     /protein_id="AAC98618.1"
                     /db_xref="GI:4071217"
                     /translation="MCKVMRIGVDARNLVNNITGISRYVMENCRELYERGHELVLYTP
                     EKPRVQLSSSEGFDFRVSNFQGPAGRLFWSQTVLPLQLRKDRIDVFWGPAHRLPALKG
                     AIPSVVSIHDLVWYYASSTMRLQGWLADRFLMKGALRNADQIVAVSYATSSAISSVFP
                     QYTSKIQMVYPGVSHFNRIGTSNILKEHRINKSYGLFIGTLEPRKNLIRLLEAYANLP
                     ECVRNNFLLVVAGGKGWNLGDIGKTINRLGIAEYTRLTGYVTDSDLNDLYRNARTLLM
                     PSLYEGFGLPIIEAQAYGVPVITSNCSSMPEVAGDAAILVDPLNVAEISNALYRIATD
```

```
                NKLWQRKSLLALQNIERFNWHNSIEKLESVFIKARESKLMTY"
repeat_region   complement(2197. .4850)
                /note="disrupted"
                /rpt_family="ISBm1"
misc_feature    complement(2210. .4820)
                /note="contains putative ISBm1 transposase coding region;
                -1 frameshifting may result in the fusion of two
                overlapping ORFs (P1 and P2 ISBm1)"
CDS             complement(join(2210. .2416,4268. .4534))
                /note="disrupted by IS711 (IS6501)"
                /codon_start=1
                /transl_table=11
                /product="P2 ISBm1"
                /protein_id="AAC98623.1"
                /db_xref="GI:4071222"
                /translation="MCPRSPARGHRKKGDGDDGGMGRSRGGLTSKIHALVDAEGRPVN
                LRLTGGQIADCAEADALTNEFGEGDILLADKGYDTNAIRARCLVPAWANIPPKSNRKD
                SFVFSRWVCRQRSLVERFFNRIKQFRDIATRYDKRPENYLAAVKLVATRIWCQSL"
repeat_region   complement(2436. .3277)
                /rpt_family="IS711"
CDS             complement(join(4393. .4652,4652. .4820))
                /codon_start=1
                /transl_table=11
                /product="P1 ISBm1"
                /protein_id="AAC98620.1"
                /db_xref="GI:4071219"
                /translation="MTRRRYELTDHEWSIISPLLPNKPRGVARVDDRRVLNGILWRFR
                TGSPWAEVPERYGPAAACYNRFVRWRKAGVWDRLFETVPKAYDGDIVMIDSTCVRVHQ
                HAATGKKGMETMVAWDVPAAGLRAKSTRSLMPKVARSIYV"
gene            5338. .6426
                /gene="gmd"
CDS             5338. .6426
                /gene="gmd"
                /function="catalyzes the conversion of GDP-mannose to
                GDP-4-keto-6-deoxymannose"
                /note="Gmd"
                /codon_start=1
                /transl_table=11
                /product="GDP-mannose dehydratase"
                /protein_id="AAC98612.1"
                /db_xref="GI:4071212"
                /translation="MKKVALITGITGQDGAYLAELLLQKGYAVHGIKRRTSQFNTARI
                DHLYHDPHEQGVDLTLHHGDLTDTSSLVRIMQLVRPDEVYNLGAQSHVAVSFEEPEYT
                ANSDALGALRLLEAIRILGMEKQTRYYQASTSELYGLVQEIPQKETTPFYPRSPYAAA
                KLYAYWITVNYREAYGIYACNGILFNHESPLRGETFVTRKITRALARIKIGNQNRLYL
                GNLDSLRDWGHARDYVEMQWLMLQQDQPEDFVIATGKQYSVREFVTLAAKDIGIEMRW
                EGSGEAEKGYDAKTGACIVEVDPRYFRPAEVETLLGDATKAHQKLGWKPKISFETLVS
                EMIAEDLAAAARENLLREHGHQFFAPKE"
gene            6434. .7537
                /gene="perA"
CDS             6434. .7537
                /gene="perA"
                /function="catalyzes the conversion of
                GDP-4-keto-6-deoxymannose to 4-NH2-4,6-dideoxymannose
                (perosamine)"
                /note="Per"
                /codon_start=1
                /transl_table=11
                /product="perosamine synthetase"
                /protein_id="AAC98613.1"
                /db_xref="GI:3170023"
                /translation="MDIPVYSPYLCGNFKKYVNECLDTGWISSRGEFISRFEDAFAQY
                VDVPSAASVANGTVALHLALDALGIGAGDEVIVPTFTYIASVNTILQTGATPVYVDSL
                ENTLQIDPEGVRLAITERTKAVMVVHLYGHPCDMDSIREICDEKSLLLVEDCAEGFGT
                KWKNSHVGTFGDVATFSFFGNKTITTGEGGMVLARNPQVMEKCRHLKSQGTSPTREYW
                HDALAYNYRMTNIQAAIGLSQIEMADEILSLKARTAASYASKLAGLPLRMHTPVGDVK
                HSYWMCSIVLDNSEHREPLRQHLRENGVDTRPFFPPAHRMPHSASTGSYPVADSLSAR
                GLNLPSFPHITDVEISFVCDLVRSYFSNHSNHI"
gene            7552. .8334
                /gene="wzm"
CDS             7552. .8334
                /gene="wzm"
                /function="predicted integral membrane domains of
                ATP-binding cassette (ABC) transporters"
                /note="Wzm"
                /codon_start=1
                /transl_table=11
                /product="ABC transporter"
                /protein_id="AAC98614.1"
                /db_xref="GI:4071213"
                /translation="MISYMANVWKVRHFWWHLSMSDLRGRFRRSSLGILWAVIQPLAL
                TLLLSFVFSKLLNQSISAYAPYILSGIIIWEYISFTVVGGSTALVQADAYIKQTRNPL
                AIYTLRNTVSGLVVLSVASISLFGWVLIMFPENFSLSWLAIPTLLPILALIVWPLATI
                VGYIGARFRDLPNALALVLQAAWFVSPVYFKESMFRQGGLNAFVDYNPIYHVMQILRA
                PVLYGEWPTATNYIWCLGVSLLLTCVAVAVGMRAEKRAIFYL"
gene            8331. .9089
                /gene="wzt"
CDS             8331. .9089
                /gene="wzt"
                /function="predicted ATPase domain of ABC transporters"
                /note="Wzt"
                /codon_start=1
                /transl_table=11
                /product="ABC transporter"
                /protein_id="AAC98615.1"
                /db_xref="GI:4071214"
                /translation="MIQPSITLSNVHLHYAASAFKERSVKTLVNALLSMRRSAGANIE
                DIHALKGISVDIARGERVALIGHNGAGKSTFLKTIAGLYPISSGTLKVTGTVRSLFDI
                GLGFEPDATGRENILYRGLLLGLTPRFMREIEDEIIEFADLGDFIDYPIKTYSAGMQV
                RLAFAISTAVDGDILLLDEVIGAGDAAFMTKAKARIMNMVEKAEIMVLASHDLANVRQ
                LCTRALVFKAGTIAFDGRVEDAISFYNSGMGAIA"
gene            9086. .9940
                /gene="wbkB"
CDS             9086. .9940
                /gene="wbkB"
                /codon_start=1
                /transl_table=11
                /product="WbkB"
                /protein_id="AAC98616.1"
                /db_xref="GI:4071215"
                /translation="MIRDSGAPRRVLWLLNHTTLRECEVPLLQSMGFEIFTPKRFPRN
                SDNRSASVSFETDESLTIPASVIDELNSYDSYQAPVNPRIDFLINYYFENAIVAYMFP
                MFTQIISRFKGRILLRAFGLTSETWTYFDFANFVAGSFFKRTWMKASNQFWFAASYSS
                LIEIEPNFIRERTVLLPVGLPERILAKQDTWRGGDKRIMFVCPDIETYPEAKAAYSES
                KHVFGDLSHVICGNQTIPVVDDDNVVDRLSAEDGGRLPIAMSCTNCSANCTNVTQVGM
                QSRVSIYA"
gene            9967. .10746
                /gene="wbkC"
CDS             9967. .10746
                /gene="wbkC"
                /function="catalyzes the conversion of
                GDP-4-NH2-4,6-dideoxymannose to
                GDP-4-formamido-4,6-dideoxymannose"
                /note="WbkC"
                /codon_start=1
                /transl_table=11
                /product="formyl transferase"
                /protein_id="AAC98617.1"
                /db_xref="GI:4071216"
                /translation="MAIAPNTRVLVAGYGLPAEFCVTTLIGMGVEIDKIAVATHREDN
                RNCGLHSMLRLRNIQFTTAAANSEEFYEFGANFDPDMIISMHYRSLIPGRFLKLAKKG
                SVNLHPSLLPAYRGTNSVAWVIINGESETGFSYHRMDENFDTGAILLQERISVEETDT
                AFSLFHRQIARAMLRLEEVILKLDQGDPGFAQLGEASYYARELPFGGVIDPRWSEVQI
                DRFIRAMFFPPFPPAVLKIDGKVYYVPSIDIYRSLMRGIPS"
repeat_region   complement(10855. .12130)
                /rpt_family="ISBm2"
misc_feature    complement(11082. .12022)
                /note="contains putative ISBm2 transposase coding region;
                -1 frameshifting may result in the fusion of two
                overlapping ORFs (P1 and P2 ISBm2)"
CDS             complement(11082. .11804)
                /codon_start=1
                /transl_table=11
                /product="P2 ISBm2"
                /protein_id="AAC98621.1"
                /db_xref="GI:4071220"
```

```
                     /translation="MVTPAASRKAVAHLMDHHQMSERRTCKAMGFCCMAIRYETRRSD
                     DHDLRERMKALAHERRCFGYRRLQVLLRREGHLVNHKWLFRLYREEKLTVRKRGGRKR
                     AIGTRAPMLIPPAANVRWLLDFVSDQLTDGRRFRVLAVVDDCTRECLPLVADTSLSGL
                     RVARELDRIIERRGKPKMIVSDNGSEFTSNAILHWVDRTKMEWHYIAPGKPIQNAFIE
                     SFNGQLRDELLNETFFSSLTHA"
     CDS             complement(11798. .12022)
                     /codon_start=1
                     /transl_table=11

  Query Match             9.6%;  Score 97.4;  DB 1;  Length 13835;
  Best Local Similarity   52.9%;  Pred. No. 4.5e-07;
  Matches  209;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy        604 GATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTC 663
              ||| | |  ||    ||||  |  | |||  |   |||   |       ||   |||||
Db       1377 GATATTGGGAAGACTATAAATCGGCTTGGCATTGCTGAGTATACGCGTTTAACTGGGTAT 1318

Qy        664 GTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCT 723
              ||  | || |  ||| ||       |||||||       |         |  |  | || |||
Db       1317 GTTACAGACAGCGATCTAAACGATTTATATAGAAATGCACGGACGCTTCTCATGCCGTCT 1258

Qy        724 TTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTA 783
              | || |||||||||||| |||||  || | ||||     ||||  ||  |||||||  |
Db       1257 CTCTACGAAGGTTTTGGCTTACCCATAATTGAGGCACAAGCCTACGGAGTTCCTGTGATT 1198

Qy        784 ACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGACCCT 843
              || ||      ||||||  |||| ||||||||||| |||||||| ||  | |||| |||
Db       1197 ACGTCAAACTGTTCATCGATGCCTGAGGTGGCTGGGGATGCAGCAATCTTGGTCGATCCT 1138

Qy        844 CTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTTCGT 903
              ||    | ||  |  ||  | || | | |||  ||  |   || ||||  | ||||  |
Db       1137 CTGAACGTAGCCGAAATATCGAATGCACTTTATAGAATCGCTACTGATAATAAACTATGG 1078

Qy        904 AAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTGGTT 963
              |||          |   |   |            || | ||||| ||||| ||     |
Db       1077 CAGCGGAAGTCATTACTTGCTTTACAAAACATTGAGCGTTTCAACTGGCATAATAGTATA 1018

Qy        964 AGTGAGATTGAAATGGTACTGACAGAGGCATGTGA 998
                  || | ||||  ||  | |   | ||  | ||
Db       1017 GAAAAGCTGGAAAGTGTTTTCATCAAAGCCAGAGA 983


Search completed: May 11, 2006, 19:14:36
Job time : 3645.05 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 07:09:59 ; Search time 460.665 Seconds
(without alignments)
14626.705 Million cell updates/sec

Title: US-09-423-093B-3_COPY_6313_7323
Perfect score: 1011
Sequence: 1 atgaataataaaaaggtttt.....catgtgatggaaataaatga 1011

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1011	100.0	12441	2	AAX06750	Aax06750 Salmonell
	2	270.6	26.8	1029	11	ACH95192	Ach95192 Klebsiell
	3	80.8	8.0	13402	14	AEA37602	Aea37602 Shigella
	4	74	7.3	1161	11	ACH95358	Ach95358 Klebsiell
	5	67.4	6.7	14309	14	AEC31432	Aec31432 E. coli t
	6	61.6	6.1	1014	8	AAL60713	Aal60713 Yersinia
	7	61.6	6.1	1182	11	ABD03868	Abd03868 Pseudomon
	8	60.6	6.0	10450	13	ADW64302	Adw64302 E. coli O
	9	59.8	5.9	430	14	ADZ71008	Adz71008 Human chr
	10	58.8	5.8	110000	2	AAV21209_03	Continuation (4 of
	11	56.6	5.6	22080	2	AAX06751	Aax06751 Salmonell
	12	55	5.4	5254	6	ABK31514	Abk31514 Signal tr
c	13	54.8	5.4	14142	6	ABA01435	Aba01435 Streptoco
	14	54.8	5.4	20555	6	ABA01443	Aba01443 Streptoco
c	15	53.4	5.3	1554	14	ADZ70999	Adz70999 Human chr
	16	52.2	5.2	16110	6	ABA01437	Aba01437 Streptoco
	17	52	5.1	16258	6	ABK40038	Abk40038 Human che
	18	52	5.1	16258	6	ABL70376	Abl70376 Chemicall
	19	51.4	5.1	5040	2	AAX20272	Aax20272 Borrelia
	20	51.2	5.1	6013	6	ABK31361	Abk31361 Signal tr
	21	51.2	5.1	6013	6	AAS61265	Aas61265 Human gen
	22	50.8	5.0	15601	14	AEA31033	Aea31033 Escherich
	23	50.4	5.0	1116	13	ADT46895	Adt46895 Bacterial
c	24	50.4	5.0	2000	11	ACL37108	Acl37108 Rice stre
	25	50.4	5.0	349980	5	AAH41226	Aah41226 Pyrococcu
	26	50.4	5.0	349980	5	AAH41225	Aah41225 Pyrococcu
c	27	50.2	5.0	562	14	ADZ71023	Adz71023 Human chr
c	28	50	4.9	9706	4	AAK86270	Aak86270 Human imm
c	29	50	4.9	41991	13	ABD32684	Abd32684 Human can
c	30	49.6	4.9	1086	14	ADZ71081	Adz71081 Human chr
	31	49.6	4.9	1173	13	ADS46350	Ads46350 Bacterial
	32	49.6	4.9	15548	6	ABL34154	Abl34154 Human imm
c	33	49.6	4.9	110000	2	AAV21209_15	Continuation (16 o
	34	49.4	4.9	883	4	AAL15210	Aal15210 Human bre
	35	49.4	4.9	960	11	ACN85231	Acn85231 Breast ca
c	36	49.4	4.9	50000	6	ABL56201	Abl56201 AmEPV gen
	37	49	4.8	8056	8	ABZ10246	Abz10246 Haematopo
	38	49	4.8	110000	6	ABA92787_1	Continuation (2 of
	39	48.6	4.8	1194	8	ACA28619	Aca28619 Prokaryot
	40	48.6	4.8	1554	14	ADZ70999	Adz70999 Human chr
	41	48.6	4.8	14657	14	AEA31026	Aea31026 Shigella
	42	48.6	4.8	18274	6	ABA01446	Aba01446 Streptoco
	43	48.2	4.8	2000	11	ACL37108	Acl37108 Rice stre
	44	48.2	4.8	15674	6	ABL32363	Abl32363 Human imm
	45	48.2	4.8	15674	6	ABL34477	Abl34477 Human met

ALIGNMENTS

RESULT 1
AAX06750
ID AAX06750 standard; DNA; 12441 BP.
XX
AC AAX06750;
XX
DT 26-APR-1999 (first entry)
XX
DE Salmonella enterica serovar C2 O antigen gene cluster.
XX
KW O antigen; diagnosis; wbaR gene; wzx gene; wbaL gene; wbaQ gene;
KW wxy gene; wbaW gene; wbaZ gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Salmonella enterica.
XX
FH Key Location/Qualifiers
FT CDS 138. .1019
FT /*tag= a
FT /note= "abe gene; encodes AAW88324"
FT CDS 1019. .2359
FT /*tag= b
FT /product= "flippase"
FT /note= "wzx gene, this region is specifically claimed in
FT Claim 9; encodes AAW88325"
FT CDS 2362. .3314
FT /*tag= c
FT /product= "abequosyl_transferase"
FT /note= "wbaR gene, this region is specifically claimed in
FT Claim 9; encodes AAW88326"
FT CDS 3361. .3975
FT /*tag= d
FT /product= "acetyl_transferase"
FT /note= "wbaL gene, this region is specifically claimed in
FT Claim 9; encodes AAW88327"
FT CDS 3977. .5020
FT /*tag= e
FT /product= "rhamnosyl_transferase"
FT /note= "wbaQ gene, this region is specifically claimed in
FT Claim 9; encodes AAW88328"
FT CDS 5114. .6313
FT /*tag= f

```
FT                   /product= "O_antigen_polymerase"
FT                   /note= "wzy gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88329"
FT   CDS             6313. .7323
FT                   /*tag= g
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88330"
FT   CDS             7310. .8467
FT                   /*tag= h
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88331"
FT   CDS             8482. .9903
FT                   /*tag= i
FT                   /note= "manC gene; encodes AAW88332"
FT   CDS             9911. .11347
FT                   /*tag= j
FT                   /note= "manB gene; encodes AAW88333"
FT   CDS             11426. .12441
FT                   /*tag= k
FT                   /note= "wbaP gene; encodes AAW88334"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88324, AAW88325, AAW88326, AAW88327, AAW88328, AAW88329,
DR   AAW88330, AAW88331, AAW88332, AAW88333, AAW88334.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 9; Fig 9; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Salmonella enterica serotype C2 (strain M67) O
CC   antigen. 11 Open reading frames within the gene cluster were identified
CC   on the basis of homology to known sequences. The genes encode proteins
CC   (see AAW88323-34) involved the synthesis of sugars present in the
CC   polysaccharide antigen, and in the transport or processing of
CC   polysaccharide or oligosaccharide units. The use of nucleic acid
CC   molecules derived from assembly and transport genes, particularly wbd
CC   (transferase), wzx (flippase) and wzy (polymerase) genes, within O
CC   antigen gene clusters improves the specificity of methods for the
CC   detection and identification of O antigens, e.g. in testing food- or
CC   faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of S. enterica
XX
SQ   Sequence 12441 BP; 3887 A; 1876 C; 2518 G; 4160 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 1011;  DB 2;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-218;
  Matches 1011;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6313 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 6372

Qy         61 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6373 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 6432

Qy        121 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6433 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 6492

Qy        181 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6493 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 6552

Qy        241 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6553 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 6612

Qy        301 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6613 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 6672

Qy        361 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6673 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 6732

Qy        421 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6733 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 6792

Qy        481 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6793 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 6852

Qy        541 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6853 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 6912

Qy        601 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6913 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 6972

Qy        661 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6973 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 7032

Qy        721 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7033 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 7092

Qy        781 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7093 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 7152

Qy        841 CCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7153 CCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTT 7212

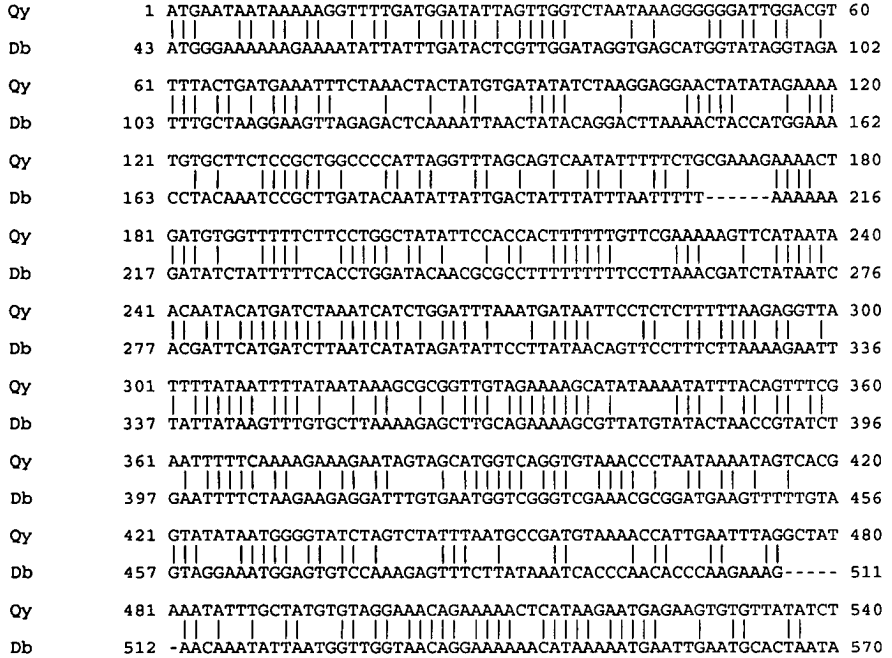
Qy        901 CGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7213 CGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTG 7272

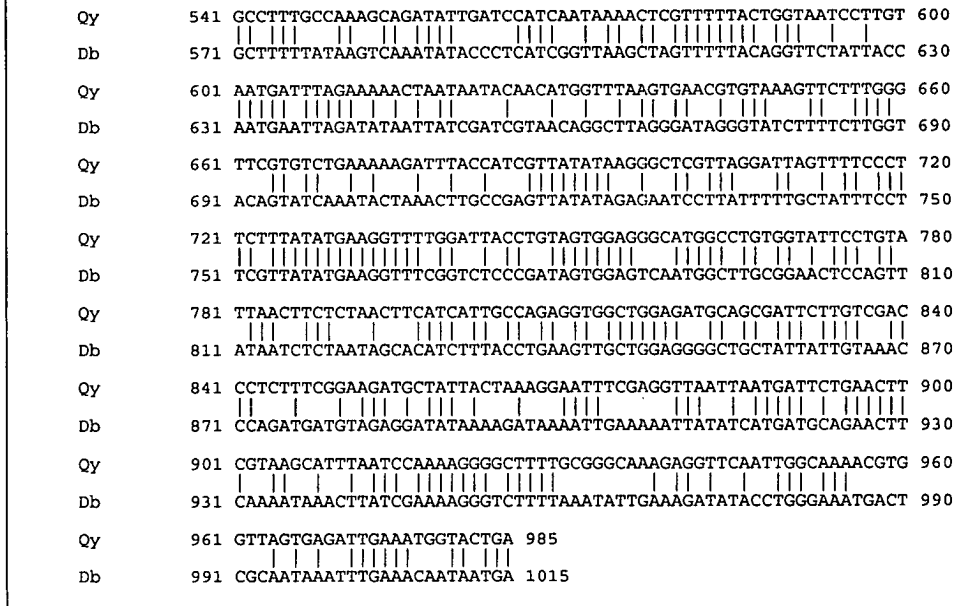
Qy        961 GTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGATGGAAATAAATGA 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7273 GTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGATGGAAATAAATGA 7323

RESULT 2
ACH95192
ID   ACH95192 standard; DNA; 1029 BP.
XX
AC   ACH95192;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Klebsiella pneumoniae polynucleotide seqid 987.
XX
KW   Recombinant expression vector; transcription regulatory element;
```

KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.
XX
OS Klebsiella pneumoniae.
XX
PN US6610836-B1.
XX
PD 26-AUG-2003.
XX
PF 27-JAN-2000; 2000US-00489039.
XX
PR 29-JAN-1999; 99US-0117747P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton GL, Osborne M;
XX
DR WPI; 2003-895346/82.
DR P-PSDB; ABO61641.
XX
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS Disclosure; SEQ ID NO 987; 932pp; English.
XX
CC The invention describes a new isolated nucleic acid encoding a Klebsiella
CC pneumoniae polypeptide. Also described are: a recombinant expression
CC vector comprising the nucleic acid, operably linked to a transcription
CC regulatory element; and a cell comprising the recombinant expression
CC vector. The nucleic acid is useful for preparing a vaccine composition
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella
CC pneumoniae polypeptide of the invention
XX
SQ Sequence 1029 BP; 368 A; 144 C; 183 G; 334 T; 0 U; 0 Other;

Query Match 26.8%; Score 270.6; DB 11; Length 1029;
Best Local Similarity 56.2%; Pred. No. 2.4e-51;
Matches 554; Conservative 0; Mismatches 419; Indels 12; Gaps 2;





RESULT 3
AEA37602
ID AEA37602 standard; DNA; 13402 BP.
XX
AC AEA37602;
XX
DT 28-JUL-2005 (first entry)
XX
DE Shigella boydii 10 O-antigen gene cluster.
XX
KW O-antigen; ds.
XX
OS Shigella boydii.
XX
PN CN1546510-A.
XX
PD 17-NOV-2004.
XX
PF 03-DEC-2003; 2003CN-01107160.
XX
PR 03-DEC-2003; 2003CN-01107160.
XX
PA (UYNA-) UNIV NANKAI.
XX
PI Feng L, Yang J, Peng X;
XX
DR WPI; 2005-215506/23.
XX
PT Nucleotide specific for Shigella boydii 10 O-antigen useful for detecting
PT and identifying Shigella boydii 10 in human body and the environment.
XX
PS Claim 1; SEQ ID NO 1; 32pp; Chinese.
XX
CC The invention relates to a nucleotide specific for Shigella boydii 10 O-
CC antigen, which is the total nucleotide sequence of the gene cluster for
CC controlling the synthesis of O-antigens in Shigella boydii 10. The
CC present sequence represents the Shigella boydii 10 O-antigen gene
CC cluster.
XX

```
SQ   Sequence 13402 BP; 4348 A; 2224 C; 2646 G; 4184 T; 0 U; 0 Other;

  Query Match             8.0%;  Score 80.8;  DB 14;  Length 13402;
  Best Local Similarity   47.3%;  Pred. No. 3.4e-08;
  Matches  315;  Conservative    0;  Mismatches  342;  Indels    9;  Gaps    2;

Qy        231 GTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTT 290
              | | || | |||| ||||||||||    || | | |  |||     |   |
Db       4554 GATAATCACAACAGTACATGATCTGTGTCCTGTTGCATTACGGAAATTATTTGGATCCAA 4613

Qy        291 TAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATT 350
                ||   ||  ||||    | |  ||          |     |  |||| ||| |
Db       4614 AAAAGCTGTAGCTTATTGGGCTTTGCTATTAATACAACTTGTTTGTTCTAATAAGATAAT 4673

Qy        351 TACAGTTTCGAATTT---TTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAA 407
              | |  |||| |  ||   | | |||||||  ||        |  |  ||| ||      |
Db       4674 TTCCATTTCAAGATTCACTGCTAAAGAAATAAATAAGTATTACTTTGGTTTATTTAAAAG 4733

Qy        408 TAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAA------A 461
              ||||  ||        |    |||||   ||   | | | ||||||    ||  |||       |
Db       4734 TAAATTATATGTCATTGGCAATGGTATAAGAAATATTTTTAATTTGGAATTAAATAGTGA 4793

Qy        462 ACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAA 521
              | ||   ||||      |  |||||    | || || || || |  ||  ||||||| ||
Db       4794 AACAGAAAATTCCCATGAAGAATATGGTATTTGCGTGGGCAATATTAAGCCTCATAAAAA 4853

Qy        522 TGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGT 581
              ||          | ||   |  | ||   | |       ||| |   |  |||| |  |  |
Db       4854 TGTTGTACCACTAATTGATTACCTTAAGACAAACATCATTAACAAAAAAATATATTTTAT 4913

Qy        582 TTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGA 641
                |    |||   | |         | |||   ||   | |   | | |||   |||
Db       4914 AGGTGAAATAAATGGGTTTTCAACGAAAATAAATAAAAGTTTGCTTAACGATTTGCCTGA 4973

Qy        642 ACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTC 701
              |  | ||   ||    |||   ||   |||  |||| ||||     |  ||        |
Db       4974 AAATATAGTTTTTACAGGGAGAGTAAGTGATCAAGAATTACATAATTATTACCGTAATGC 5033

Qy        702 GTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCAT 761
                ||  | |||| || ||||| |||| |||||  || ||| |  | ||||  ||
Db       5034 CAGTTTTTTTATATTCCCCTCATTATACGAAGGCTTTGGTTTGCCTTTGCTAGAGGCTAT 5093

Qy        762 GGCCTGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGA 821
              |   | | |  | ||| | ||  | ||    |||    || || ||||| ||
Db       5094 GAATTATAGGTTGCCTATTTTTGCATCAAATATTGATGTATTTAAGGAAGTTGCTGGTGA 5153

Qy        822 TGCAGCGATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTT 881
              ||   | |  | | || ||   | |    ||||||  || |  |  |||||    |  ||
Db       5154 AACAATTAATTATTTTGATCCATATACATTTGATGCTCTTGCAGATAGAATTAACAAATT 5213

Qy        882 AATTAA 887
              ||  ||
Db       5214 AAAAAA 5219


RESULT 4
ACH95358
ID   ACH95358 standard; DNA; 1161 BP.
XX
AC   ACH95358;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Klebsiella pneumoniae polynucleotide seqid 1153.
XX
KW   Recombinant expression vector; transcription regulatory element;
KW   Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.
XX
OS   Klebsiella pneumoniae.
XX
PN   US6610836-B1.
XX
PD   26-AUG-2003.
XX
PF   27-JAN-2000; 2000US-00489039.
XX
PR   29-JAN-1999;   99US-0117747P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Breton GL,  Osborne M;
XX
DR   WPI; 2003-895346/82.
DR   P-PSDB; ABO61807.
XX
PT   New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT   preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS   Disclosure; SEQ ID NO 1153; 932pp; English.
XX
CC   The invention describes a new isolated nucleic acid encoding a Klebsiella
CC   pneumoniae polypeptide. Also described are: a recombinant expression
CC   vector comprising the nucleic acid, operably linked to a transcription
CC   regulatory element; and a cell comprising the recombinant expression
CC   vector. The nucleic acid is useful for preparing a vaccine composition
CC   against Klebsiella pneumoniae. This sequence encodes a Klebsiella
CC   pneumoniae polypeptide of the invention
XX
SQ   Sequence 1161 BP; 398 A; 166 C; 204 G; 393 T; 0 U; 0 Other;

  Query Match             7.3%;  Score 74;  DB 11;  Length 1161;
  Best Local Similarity   52.2%;  Pred. No. 7.1e-07;
  Matches  164;  Conservative    0;  Mismatches  150;  Indels    0;  Gaps    0;

Qy        611 AAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTG 670
              ||  || | ||||| ||    ||      | ||||  | |  ||| | |  |  ||    |
Db       767 AATCACCATTAATAAAATTACGTAATGATAAACGCTTTATATTCATCGATTATGTCGAAG 826

Qy        671 AAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTTATATG 730
              ||      |||   | | | |||||||    | |||  | ||  |||  || |||||| |||
Db       827 GAAGTACATTAGATTGGCTTTATAAGAATGCATTTTGTTTTATTTATCCCTCTTTAAATG 886

Qy        731 AAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTAACTTCTC 790
              |||| | ||| |  ||      |  ||| ||||   |  ||||  || ||| | | |||
Db       887 AAGGATATGGTTATCCGCCGTTACAGGCCATGAGGTACGGTACACCCGTAATTGCATCTT 946

Qy        791 TAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGACCCTCTTTCGG 850
              || |||    | ||||| ||   || ||||    |   | | || ||
Db       947 GCACATCAAGTATTCCAGAAGTATGTGCAGATAATGTACTATATTTTGATCCAAGGAATA 1006

Qy        851 AAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTTCGTAAGCATT 910
              |   |  ||    ||  |||||  |  |||  ||  |||  |||||||  |||  |
Db       1007 TAAGCGAAATCGAAAATCGAATTCTGCAATTATCTACAGATGGTGAACTTTATAATAAAC 1066

Qy        911 TAATCCAAAAGGGG 924
              ||       ||||||
Db       1067 TAGCTTTGAAGGGG 1080


RESULT 5
AEC31432
ID   AEC31432 standard; DNA; 14309 BP.
XX
AC   AEC31432;
XX
DT   20-OCT-2005  (first entry)
XX
DE   E. coli type O43 O-antigen gene cluster.
XX
KW   O-antigen; gene expression; DNA detection; gene; ds.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1083. .2396
FT                   /*tag=  a
```

```
FT                   /note= "Orf1"
FT   CDS             2412. .3155
FT                   /*tag=  b
FT                   /note= "Orf2. Specifically claimed in Claim 3"
FT   CDS             3152. .3673
FT                   /*tag=  c
FT                   /note= "Orf3. Specifically claimed in Claim 3"
FT   CDS             3652. .4740
FT                   /*tag=  d
FT                   /note= "Orf4. Specifically claimed in Claim 3"
FT   CDS             4712. .5938
FT                   /*tag=  e
FT                   /note= "Orf5"
FT   CDS             5935. .6984
FT                   /*tag=  f
FT                   /note= "Orf6. Specifically claimed in Claim 3"
FT   CDS             6981. .8099
FT                   /*tag=  g
FT                   /note= "Orf7"
FT   CDS             8103. .9068
FT                   /*tag=  h
FT                   /note= "Orf8"
FT   CDS             9071. .9526
FT                   /*tag=  i
FT                   /note= "Orf9"
FT   CDS             9538. .10941
FT                   /*tag=  j
FT                   /note= "Orf10"
FT   CDS             11007. .11753
FT                   /*tag=  k
FT                   /note= "Orf11. Specifically claimed in Claim 3"
FT   CDS             11778. .13148
FT                   /*tag=  l
FT                   /note= "Orf12"
XX
PN   CN1563043-A.
XX
PD   12-JAN-2005.
XX
PF   19-APR-2004; 2004CN-00019028.
XX
PR   19-APR-2004; 2004CN-00019028.
XX
PA   (TIAN-) TIANJIN BIOCHIP TECH CO LTD.
XX
PI   Wang L,  Yang J,  Feng L;
XX
DR   WPI; 2005-307118/32.
XX
PT   Nucleotide peculiar to O-antigen of Escherichia coli O43.
XX
PS   Claim 1; SEQ ID NO 1; 35pp; Chinese.
XX
CC   The invention relates to a nucleotide sequence (SEQ ID No:1) which is
CC   specific to the O-antigen of Escherichia coli O43. The nucleotide
CC   sequence represents the complete sequence (14309bp) of a gene cluster for
CC   controlling the synthesis of O-antigen in Escherichia coli O43. The
CC   nucleotide sequence can have one or several inserted, deleted or
CC   substituted bases, whilst still retaining its function. Also described is
CC   a PCR primer for the glycosyltransferase (galF) gene and a PCR primer for
CC   the oligosaccharide unit treatment (gnd) gene, both of which are
CC   contained within the E. coli type O43 O-antigen gene cluster. The
CC   invention also discloses a method for detecting and identifying E. coli
CC   type O43 in humans. This sequence represents the E. coli type O43 O-
CC   antigen gene cluster of the invention.
XX
SQ   Sequence 14309 BP; 4529 A; 2267 C; 2972 G; 4541 T; 0 U; 0 Other;

  Query Match             6.7%;  Score 67.4;  DB 14;  Length 14309;
  Best Local Similarity   50.3%;  Pred. No. 3.7e-05;
  Matches  220;  Conservative    0;  Mismatches  211;  Indels    6;  Gaps    2;

Qy        571 ATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACAT 630
              || || || ||  ||  ||||     |  |  |||||  || || ||| |||   | ||
Db       6547 ATTAATCTTGTGATTGTTGGTGGAGGTGATCTGGATTTTCAAGAATTAAAAATGTTAAAT 6606

Qy        631 GG---TTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCG 687
               |   |||||  | ||||    |        || |  ||   |||  |  | |||
Db       6607 AGATATTTAAAAGGACGTTATTATAAAAAAGGTTATGTAGATGACCAGCAATTAAATAAT 6666

Qy        688 TTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCT 747
               | || |   || | |||   ||| |||  || ||   ||||||||| |||||  | |||
Db       6667 CTTTACACCTGCGCTTTATTTTTATTTTATCCATCCCAATATGAAGGGTTTGGTATTCCT 6726

Qy        748 GTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCA 807
              ||   | || |  |||    | |||   |||||||  |  |  ||      ||| ||  |   |
Db       6727 GTGTTAGAAGCAATGAAGTCTGGGTGTCCTGTAATTGCAACTGATTGTTCTTCTGTAAGA 6786

Qy        808 GAGGTGGCTGGAGATGCAGCGATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAA 867
              |||  | ||   ||      ||  | || | | |        | |  ||     || ||||
Db       6787 GAGATAGCATCAGGATATGCTCTGCTAGCCAAAGAGGCAACTGTTGAACAATTTTCTAAG 6846

Qy        868 GGAATTTCGAGGTTAATTAATGATTCTGAACTTCGTAAGCATTTAATCCAAAAGGGGCTT 927
               | |||      ||| |  ||||   |    || |  || |  | ||  | || || ||
Db       6847 TGTATTAATTTATTATTAGATGAAAGT---TTTAGAGAGAAAATGATATATAATGGTCTA 6903

Qy        928 TTGCGGGCAAAGAGGTTCAATTGGCAAAACGTGGTTAGTGAGATTGAAATGGTACTGACA 987
                 || ||||  |||| |||| | ||      ||  ||| | ||  |   |
Db       6904 CATTATGCCAAGAAATTCACTTGGGATAAATGCTTTTCTGATACTGTTAGTTTTTATGAT 6963

Qy        988 GAGGCATGTGATGGAAA 1004
              ||||   |  ||  |||
Db       6964 GAGGTGAGATATTTAAA 6980


RESULT 6
AAL60713
ID   AAL60713 standard; DNA; 1014 BP.
XX
AC   AAL60713;
XX
DT   03-SEP-2003  (first entry)
XX
DE   Yersinia pestis mutant O-antigen polymerase (1B3) gene.
XX
KW   Infection; tuberculosis; veterinary treatment; antimicrobial drug;
KW   vaccine; O-antigen polymerase; enzyme; mutant; gene; ds.
XX
OS   Yersinia pestis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1014
FT                   /*tag=  a
FT                   /product= "Yersinia pestis mutant O-antigen polymerase"
XX
PN   WO2003044047-A2.
XX
PD   30-MAY-2003.
XX
PF   18-NOV-2002; 2002WO-GB005212.
XX
PR   19-NOV-2001; 2001GB-00027657.
XX
PA   (UNLO ) LONDON SCHOOL HYGIENE & TROPICAL MEDICIN.
XX
PI   Karlyshev AV,  Wren BW;
XX
DR   WPI; 2003-457588/43.
DR   P-PSDB; AAO30098.
XX
PT   New virulence peptides encoded by a gene of Yersinia pseudotuberculosis,
PT   useful for treating or preventing a condition associated with infection
PT   by Yersinia, e.g. tuberculosis, particularly in veterinary treatment.
XX
PS   Claim 1; Page 66-68; 130pp; English.
```

XX
CC The present invention relates to virulence genes of Yersinia species and
CC proteins encoded by such genes. Sequences of the invention are useful in
CC the manufacture of a medicament for treating or preventing a condition
CC associated with infection by Yersinia e.g. tuberculosis particularly in
CC veterinary treatment. They are useful in screening assays for identifying
CC antimicrobial drugs. They are also used to prepare vaccines. The present
CC sequence is Yersinia pestis mutant O-antigen polymerase gene
XX
SQ Sequence 1014 BP; 320 A; 154 C; 208 G; 332 T; 0 U; 0 Other;

Query Match 6.1%; Score 61.6; DB 8; Length 1014;
Best Local Similarity 49.7%; Pred. No. 0.00044;
Matches 157; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

```
Qy        521 ATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCG 580
              || |||| | ||  ||  | |     || | | || | |    ||        |  |   |
Db        542 ATAAGAATTTTGACATTGCAGTGAAAGCAATATCAAAAACACCTCACCTCGAATTATCAG 601

Qy        581 TTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTG 640
              || |   |||     | |  | |  |  ||||     | || |  | ||   ||||  ||
Db        602 TTGTAGGTGGAGGGGCATTCACTAGTAAAGAACTGTCACTACTGAATCACTATTTACCTG 661

Qy        641 AACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCT 700
              ||      |  || ||   | |  |||  | | |||      |  ||||||
Db        662 GTCGCTACCATGGATTAGGTCGCCTTAGTGATGAGGCTTTGAATGAGGCATATAATTCAG 721

Qy        701 CGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCA 760
              | |  |    |  ||| ||| |||   ||||| ||||||||  | ||  || | || |  |
Db        722 CTTATGCGCTGCTTTACCCATCTAGCTATGAGGGTTTTGGTATTCCAATATTAGAAGCGA 781

Qy        761 TGGCCTGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAG 820
              ||      ||   ||| ||| || ||  |      || ||  | || |||||  || || |
Db        782 TGAGTGCAGGATGTCCCGTAATATCTGTTAATGTATCTTCTATACCTGAGGTCGCAGGTG 841

Qy        821 ATGCAGCGATTCTTGT 836
              |||| || ||  | ||
Db        842 ATGCCGCTATATTAGT 857
```

RESULT 7
ABD03868
ID ABD03868 standard; DNA; 1182 BP.
XX
AC ABD03868;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polynucleotide #2472.
XX
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US6551795-B1.
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
DR WPI; 2003-615309/58.
DR P-PSDB; ABO70297.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 2472; 455pp; English.
XX
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01397-
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 1182 BP; 163 A; 392 C; 417 G; 210 T; 0 U; 0 Other;

Query Match 6.1%; Score 61.6; DB 11; Length 1182;
Best Local Similarity 50.0%; Pred. No. 0.00046;
Matches 154; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

```
Qy        646 GTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTA 705
              ||   | |  | || | | || | ||      |   |   |||     |||| |
Db        823 GTGCGGCTGCTCGGCTACCTGCCGGACGAGCGGGTCGCCGAGTTGATGTCGGGCGCTCGG 882

Qy        706 GGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCC 765
              |   | || ||||| ||  | || |||||  || ||  | |||||  |||| | ||||||||
Db        883 GCGCTGGTCTTCCCCTCGATCTACGAAGGCTTCGGCCTGCCTGTGCTGGAAGCCATGGCC 942

Qy        766 TGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCA 825
              | || |  || ||||      | | |   ||  || |||| ||||||||| || || ||
Db        943 AGCGGCACGCCAGTATTGCTCACGCGACTCTCCGCAATGCCGGAGGTGGCCGGCGAGGCG 1002

Qy        826 GCGATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATT 885
              | |     | | ||||  |   |   |||  |  || |     |     | | |
Db       1003 GGCAGCTATATTGACCCGGATGATGCCAATGGGCTATCTGACATGCTGGTCCGCATGACC 1062

Qy        886 AATGATTCTGAACTTCGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTC 945
              | ||| | |     |  ||   |    ||  || ||    ||||  ||  |||
Db       1063 GAGGATACCGTCTACTGGGAGTGCTGCCGGGAAGCCGGCCTGCGACGGGCGGAGCAGTTT 1122

Qy        946 AATTGGCA 953
                |||  |
Db       1123 TCCTGGAA 1130
```

RESULT 8
ADW64302
ID ADW64302 standard; DNA; 10450 BP.
XX
AC ADW64302;
XX
DT 24-MAR-2005 (first entry)
XX
DE E. coli Ogawa gene cluster DNA.
XX
KW Ogawa; glycosyltransferase; oligosaccharide unit processing; detection;
KW ds; gene.
XX
OS Escherichia coli.
XX
PN CN1453358-A.
XX
PD 05-NOV-2003.
XX
PF 27-MAY-2003; 2003CN-00129935.
XX

PR 27-MAY-2003; 2003CN-00129935.
XX
PA (UYNA-) UNIV NANKAN.
XX
PI Wang L, Yang J, Feng L;
XX
DR WPI; 2004-157424/16.
XX
PT Nucleotide specific to Ogawa of colibacillus 0-59 for detecting and
PT identifying colibacillus 059 in human body and in environment.
XX
PS Claim 1; SEQ ID NO 1; 31pp; Chinese.
XX
CC The invention relates to a nucleotide sequence of a gene cluster which
CC controls the synthesis of Ogawa in colibacillus 059. Also claimed are
CC oligonucleotides of glycosyltransferase gene and oligosaccharide unit
CC processing gene in Ogawa gene cluster originated from colibacillus 059.
CC The oligonucleotide is useful for detecting and identifying colibacillus
CC 059 in human body and in environment. The present sequence represents the
CC Ogawa gene cluster.
XX
SQ Sequence 10450 BP; 3198 A; 1739 C; 2162 G; 3350 T; 0 U; 1 Other;

Query Match 6.0%; Score 60.6; DB 13; Length 10450;
Best Local Similarity 54.8%; Pred. No. 0.0012;
Matches 120; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Qy 604 GATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTC 663
Db 4404 GAGTTAGATAAATTAGTAGAGAAGGGGGTGGCAAAAGGTATTATCAGAAAATTAGGTTAT 4463

Qy 664 GTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCT 723
Db 4464 GTATCTGATCCGGATTTAGCGGTTCTTTATTCAGGAGCTCAACTTTTTGTTTTAGCATCT 4523

Qy 724 TTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTA 783
Db 4524 ATTTATGAAGGTTTTGGTATGCCGATTCTGGAGGCTCAAGCCAGCGGGTGTCCTGTGCTT 4583

Qy 784 ACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGAT 822
Db 4584 ATATCCAGAATAAAATCTATGATAGAGGCCGCAGGTGAT 4622

RESULT 9
ADZ71008
ID ADZ71008 standard; DNA; 430 BP.
XX
AC ADZ71008;
XX
DT 14-JUL-2005 (first entry)
XX
DE Human chromosome 2 contig DNA SEQ ID NO 132.
XX
KW ds; matrix attachment region; MAR; protein production.
XX
OS Homo sapiens.
XX
PN WO2005040377-A2.
XX
PD 06-MAY-2005.
XX
PF 22-OCT-2004; 2004WO-EP011974.
XX
PR 24-OCT-2003; 2003US-0513574P.
PR 06-FEB-2004; 2004EP-00002722.
XX
PA (SELE-) SELEXIS SA.
XX
PI Mermod N, Girod PA, Bucher P, Nguyen D, Calabrese D, Saugy D;
PI Puttini S;
XX
DR WPI; 2005-333507/34.
XX
PT New purified and isolated DNA sequence having protein production
PT increasing activity comprises a bent DNA element and a binding site for a
PT DNA binding protein, useful for increasing protein production activity in
PT eukaryotic host cell.
XX
PS Disclosure; SEQ ID NO 132; 282pp; English.
XX
CC The invention relates to a purified and isolated DNA sequence having
CC protein production increasing activity comprising at least one bent DNA
CC element, and at least one binding site for a DNA binding protein. The
CC purified and isolated DNA sequence comprising a first and a second
CC isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC MAR nucleotide sequence selected from a purified and isolated DNA
CC sequence above, a purified and isolated MAR DNA above, a purified and
CC isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC sequence complementary to it, its molecular chimera, or its combinations
CC and variants, is useful for increasing protein production activity in a
CC eukaryotic host cell. The present sequence represents a human chromosome
CC 2 contig DNA.
XX
SQ Sequence 430 BP; 197 A; 7 C; 7 G; 219 T; 0 U; 0 Other;

Query Match 5.9%; Score 59.8; DB 14; Length 430;
Best Local Similarity 47.3%; Pred. No. 0.00094;
Matches 181; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

Qy 68 ATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAATGTGCTT 127
Db 48 ATAATATATAATATATAATATATAAACTATATATTAATATAATATATATAAACTATTATA 107

Qy 128 CTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACTGATGTGG 187
Db 108 TAAATACATATTAAATATATTATATTTTAATATTTATATATTAAATATAATATATATTT 167

Qy 188 TTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATAACAATAC 247
Db 168 AATATTTATATATTAAATATATAATATATTTAATATTTATATAATATATAGCATATTTTA 227

Qy 248 ATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTATTTTATA 307
Db 228 TATTTATATTATATATAACATTTTATATTTATATTTATATTTATATATATTTAATTTATA 287

Qy 308 ATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCGAATTTTT 367
Db 288 TTTATATTATATTTATATTTATATTATATATAACATAATTATATATATTTTCATATTGTA 347

Qy 368 CAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACGGTATATA 427
Db 348 TATAATAAAGAAATGTATATTTGTTATATATAATATATATTATATAATTTATTATATATT 407

Qy 428 ATGGGGTATCTAGTCTATTTAAT 450
Db 408 ATATAATATATATTATATAATAT 430

RESULT 10
AAV21209_03
Continuation (4 of 17) of AAV21209 from base 300001 (Methanococcus jannaschii circular ch
WP Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209

WP	Fragment Name	Begin	End
WP	AAV21209_00	1	110000
WP	AAV21209_01	100001	210000
WP	AAV21209_02	200001	310000
WP	AAV21209_03	300001	410000
WP	AAV21209_04	400001	510000
WP	AAV21209_05	500001	610000
WP	AAV21209_06	600001	710000
WP	AAV21209_07	700001	810000
WP	AAV21209_08	800001	910000
WP	AAV21209_09	900001	1010000
WP	AAV21209_10	1000001	1110000
WP	AAV21209_11	1100001	1210000

```
WP      AAV21209_12       1200001   1310000
WP      AAV21209_13       1300001   1410000
WP      AAV21209_14       1400001   1510000
WP      AAV21209_15       1500001   1610000
WP      AAV21209_16       1600001   1664976

  Query Match             5.8%;  Score 58.8;  DB 2;  Length 110000;
  Best Local Similarity   48.7%;  Pred. No. 0.005;
  Matches  224;  Conservative    0;  Mismatches  227;  Indels    9;  Gaps    2;

Qy        546 TGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGA 605
              ||| | |  |||||| |||        | |||||| |      ||| ||     |  ||  |
Db      55500 TGCAATAAAAGATATAGATTTTAATTTTAAACTCATAGGGGATGGGAAGTTATACAAAAA 55559

Qy        606 TTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGT 665
              ||||  ||  |  |  |  || ||  ||||||    | | |  |  |||| ||
Db      55560 AATAGAGAACTTTGTTGTTAAAAATAATTTAAG---CCATATTGAACTCTTAGGAAGAAA 55616

Qy        666 GTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTT 725
              |    | ||  || | || ||    | |   |        |||||  | |||||
Db      55617 AAGTTTTGATGAAGTAGCTTCATTTATGAGGAAGTGTAGTTTTTTAGTAGTTCCTTCAAG 55676

Qy        726 ATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTAAC 785
              |  |||||||||||||| |    |  || || || ||||| ||   ||  |||||| || |
Db      55677 AAGTGAAGGTTTTGGAATGGTGGCTGTTGAAGGAATGGCTTGCTCTAAGCCTGTAATAGC 55736

Qy        786 TTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGA------TTCTTGTCGA 839
              |     |     |||  ||| | | |   |||| |   |     |  | |  ||
Db      55737 CACAAGGGTTGGGGGGTTGGGGGAGATTGTTATTGATGGATATAACGGACTATTGGCTGA 55796

Qy        840 CCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACT 899
                  | ||| | | |   ||  ||||  |  |||||| ||| ||   |||||
Db      55797 GAAAAATAACCCAAATGATTTAAAAGAAAAAATTCTGGAGTTAATAAATAATGAAGAACT 55856

Qy        900 TCGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGT 959
              | ||   |||     |||| ||           ||||  |  ||   |||| ||||
Db      55857 AAGAAAAACTTTGGGGGAAAATGGAAAAGAATTTTCAAAAAAATTTTCTTGGGAAAAATG 55916

Qy        960 GGTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGAT 999
              || |  |  ||  ||  ||      ||||  | | |||
Db      55917 TGTAATGGGTGTTAGAAAAGTGTATGAAGAGCTAAGCGAT 55956


RESULT 11
AAX06751
ID   AAX06751 standard; DNA; 22080 BP.
XX
AC   AAX06751;
XX
DT   26-APR-1999  (first entry)
XX
DE   Salmonella enterica serovar B O antigen gene cluster.
XX
KW   O antigen; diagnosis; wzx gene; wbaV gene; transferase; polymerase;
KW   flippase; diarrhoea; haemorrhagic colitis; ds.
XX
OS   Salmonella enterica.
XX
FH   Key             Location/Qualifiers
FT   CDS             4099. .5184
FT                   /*tag=  a
FT                   /product= "TDP-rhamnose_pathway_protein"
FT                   /note= "rmlB gene; encodes AAW88335"
FT   CDS             5184. .6083
FT                   /*tag=  b
FT                   /product= "TDP-rhamnose_pathway_protein"
FT                   /note= "rmlD gene; encodes AAW88336"
FT   CDS             6131. .7009
FT                   /*tag=  c
FT                   /product= "TDP-rhamnose_pathway_protein"
FT                   /note= "rmlA gene; encodes AAW88337"
FT   CDS             7010. .7561
FT                   /*tag=  d
FT                   /product= "TDP-rhamnose_pathway_protein"
FT                   /note= "rmlC gene; encodes AAW88338"
FT   CDS             7567. .8559
FT                   /*tag=  e
FT                   /product= "CDP-adequose_pathway_protein"
FT                   /note= "ddhD gene; encodes AAW88339"
FT   CDS             8556. .9329
FT                   /*tag=  f
FT                   /product= "CDP-adequose_pathway_protein"
FT                   /note= "ddhA gene; encodes AAW88340"
FT   CDS             9334. .10413
FT                   /*tag=  g
FT                   /product= "CDP-adequose_pathway_protein"
FT                   /note= "ddhB gene; encodes AAW88341"
FT   CDS             10440. .11753
FT                   /*tag=  h
FT                   /product= "CDP-adequose_pathway_protein"
FT                   /note= "ddhC gene; encodes AAW88342"
FT   CDS             11781. .12680
FT                   /*tag=  i
FT                   /product= "CDP-adequose_pathway_protein"
FT                   /note= "abe gene; encodes AAW88343"
FT   CDS             12762. .14054
FT                   /*tag=  j
FT                   /product= "flippase"
FT                   /note= "wzx gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88344"
FT   CDS             14059. .15060
FT                   /*tag=  k
FT                   /product= "abequosyl_transferase"
FT                   /note= "wbaV gene, this region is specifically claimed in
FT                   Claim 9; ; encodes AAW88345"
FT   CDS             15379. .16440
FT                   /*tag=  l
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaU gene; encodes AAW88346"
FT   CDS             16441. .17385
FT                   /*tag=  m
FT                   /product= "rhamnosyl_transferase"
FT                   /note= "wbaN gene; encodes AAW88347"
FT   CDS             17386. .18825
FT                   /*tag=  n
FT                   /product= "GDP_mannose_pathway_protein"
FT                   /note= "manC gene; encodes AAW88348"
FT   CDS             18812. .20245
FT                   /*tag=  o
FT                   /product= "GDP_mannose_pathway_protein"
FT                   /note= "mabB gene; encodes AAW88349"
FT   CDS             20317. .21747
FT                   /*tag=  h
FT                   /product= "galactosyl_transferase"
FT                   /note= "wbaP gene; encodes AAW88350"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88335, AAW88336, AAW88337, AAW88338, AAW88339, AAW88340,
DR   AAW88341, AAW88342, AAW88343, AAW88344, AAW88345, AAW88346, AAW88347,
DR   AAW88348, AAW88349, AAW88350.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
```

PT samples.
XX
PS Claim 9; Fig 10; 165pp; English.
XX
CC This is the nucleotide sequence of a gene cluster involved in the
CC biosynthesis of the Salmonella enterica serotype B (strain LT2) O
CC antigen. 16 Open reading frames within the gene cluster were identified
CC on the basis of homology to known sequences. The genes encode proteins
CC (see AAW88335-50) involved the synthesis of sugars present in the
CC polysaccharide antigen, and in the transport or processing of
CC polysaccharide or oligosaccharide units. The use of nucleic acid
CC molecules derived from assembly and transport genes, particularly wbd
CC (transferase), wzx (flippase) and wzy (polymerase) genes, within O
CC antigen gene clusters improves the specificity of methods for the
CC detection and identification of O antigens, e.g. in testing food- or
CC faecal-derived samples, or samples from patients. The O antigen is a
CC major virulence factor of S. enterica
XX
SQ Sequence 22080 BP; 6284 A; 4028 C; 5152 G; 6616 T; 0 U; 0 Other;

Query Match 5.6%; Score 56.6; DB 2; Length 22080;
Best Local Similarity 46.8%; Pred. No. 0.011;
Matches 296; Conservative 0; Mismatches 319; Indels 18; Gaps 3;

```
Qy      225 GAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTC 284
            || |||    | |||||  |||||||| |||  | | |  |||     |||| ||
Db    15690 GAGAAACCAAACGATAACCGTACATGATATAAGACCTTTTTATTATCCAGATAGTT---- 15745

Qy      285 TCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAA 344
              || || |||   | | || |   ||| || ||||         ||    |   | |
Db    15746 --TTATTCAGAAAGTGTATTTTCGCTTTTTATTAAAAATGTCCGTTAAGCGATGTAAGCA 15803

Qy      345 AATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCC 404
            | || || || ||  ||      ||||| || || |   |   |    ||||  |
Db    15804 TGTTTTAACGGTATCTTATACCGTTAAAGATAGCATTGCTAAAACTTATAATGTAGATAG 15863

Qy      405 TAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACC 464
            | | ||||||       | |||||        |||| |   |||  | ||| | | ||
Db    15864 TGAGAAAATATCAGTAATTTATAA---------TAGTGTTAATAAATCTGATTTTATACA 15914

Qy      465 ATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGA 524
            |   ||   |    | || || ||  ||  ||| ||    ||      | ||||| |||
Db    15915 AAAAAAAGAAAAAGAGAATTACTTTTTAGCTGTTGGTGCAAGTTGGCCACATAAAAATAT 15974

Qy      525 GAAGTGTGTTAT---ATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGT 581
              | |   | ||   |  |    |    | ||   | ||   | ||| |   |
Db    15975 TCATTCATTCATAAAAAATAAAAAAGTTTGGTCTGACTCTTATAATTTAATTATTGTATG 16034

Qy      582 TTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGA 641
            |  |  | |  |    |    |     | |  || || |  |  |  | |  |   ||| ||
Db    16035 TGGTCGTACTGACTATGCAATGTCTCTCCAACAAATGGTCGTTGATCTGGAACTAAAAGA 16094

Qy      642 ACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTC 701
              || |  || ||     || ||    || || |||      ||||||       |
Db    16095 TAAAGTGACTTTTTTACATGAAGTCTCATTTAATGAATTAAAGATTTTATATTCTAAAGC 16154

Qy      702 GTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCAT 761
            |  |   | ||||  || ||| |  |||| ||||||||  |||||    | || |  ||
Db    16155 CTACGCGCTTGTTTATCCATCTATTGATGAGGGTTTTGGTATACCTCCTATTGAAGCGAT 16214

Qy      762 GGCCTGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGA 821
            ||| |    || ||| ||  ||    ||     |   ||  ||| |  || |||       |
Db    16215 GGCATCAAATACTCCAGTTATAGTGTCCGATATACCAGTATTTCATGAAGTGTTAACCAA 16274

Qy      822 TGCAGCGATTCTTGTCGACCCTCTTTCGGAAGA 854
            ||  ||  |   |||  | ||   |   ||| |
Db    16275 TGGTGCATTATATGTGAATCCGGATGATGAAAA 16307
```

RESULT 12
ABK31514
ID ABK31514 standard; DNA; 5254 BP.
XX
AC ABK31514;
XX
DT 23-APR-2002 (first entry)
XX
DE Signal transduction associated gene modified DNA #179.
XX
KW Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200926-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007472.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-147896/19.
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction.
XX
PS Claim 1; SEQ ID NO 357; 24pp; English.
XX
CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC the cytosine methylation state (CpG islands) of these genes, and a method
CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC genes associated with signal transduction. The genomic DNA can be
CC obtained from cells or cellular components which contain DNA, e.g. cell
CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
CC brain, heart, prostate, lung, breast or liver, histologic object slides,
CC and all their possible combinations. The sequences of the invention are
CC useful for the diagnosis and therapy of diseases associated with signal
CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC chemically pretreated genomic DNA sequences of different genes associated
CC with signal transduction, or their complementary sequences. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office
XX
SQ Sequence 5254 BP; 1636 A; 50 C; 1034 G; 2534 T; 0 U; 0 Other;

Query Match 5.4%; Score 55; DB 6; Length 5254;
Best Local Similarity 45.5%; Pred. No. 0.019;
Matches 276; Conservative 0; Mismatches 325; Indels 6; Gaps 2;

```
Qy      198 TGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAA 257
            ||| |||  |        |||||||||      |  ||    |||| |  | |  |  | |
Db    3108 TGGTTATTATAGTTTTTTTTTTTGTTTAGTTAGTTTGTATATAAGAGAAAAATAAATGGA 3167

Qy      258 TCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAAT 317
            | || |  | ||      ||  |   | | ||||||  |||  |||||     ||| | |
Db    3168 TGATGTTTAGTTTTTGAATTTTAATTTTTTTTTTA--AGGGAATTTTTATTGTTTTTTAG 3225

Qy      318 AAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAG 377
              ||     |  |  |||   || | ||  || | ||||   |  ||||   |||  |
Db    3226 TTAGGTTTATATAATTAAATATTAGATAAAGTTAAAAGTTGTTAGATTTTGGTAAGGTAA 3285

Qy      378 AATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATC 437
```

```
Db       3286 AAGAAGAGGTTGTTTTTGTTTTTAGAATATTTAGAGTATTGTGGTATATAATGAAATAAT 3345

Qy        438 TAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGT 497

Db       3346 AGGATTGTTTTGATATTTTTTTTTTGGGGTTTATTGTTTTTTTTTATTATGTTTGGTTTGT 3405

Qy        498 AGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGA 557

Db       3406 ATTTTGAAATTTAATATAAATTATAGGAAAATTTTTTATGGAGTTAATATTTTTTATAAG 3465

Qy        558 TATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACT 617

Db       3466 TTTTGTTAAATATTTTATTTTAAATTTGTTTAGTATGGTATATAGATAAAAGTAAGAGGT 3525

Qy        618 AATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGA 677

Db       3526 TTATATATAAAAATGTATAAGTT----TGTTTTTTTTGTAATAGTGTAAAATTAAAAGGA 3581

Qy        678 TTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTT 737

Db       3582 TTTTAGAAAGGTAATTAATAGTTTGGTAGTATGTGTATTTTTTTTTTTTTAATTAGATAT 3641

Qy        738 TGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTAACTTCTCTAACTTC 797

Db       3642 TATTAGATTTTAAGAATTTGTAATGAAATTTAGGTTTATTGTATAATTTTTTATAAATAA 3701

Qy        798 ATCATTG 804

Db       3702 GTTTTTG 3708
```

RESULT 13
ABA01435/c
ID ABA01435 standard; DNA; 14142 BP.
XX
AC ABA01435;
XX
DT 21-FEB-2002 (first entry)
XX
DE Streptococcus thermophilus eps5 operon #1.
XX
KW Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
KW ds.
XX
OS Streptococcus thermophilus.
XX
PN WO200179500-A2.
XX
PD 25-OCT-2001.
XX
PF 18-APR-2001; 2001WO-FR001199.
XX
PR 18-APR-2000; 2000FR-00004972.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
PA (DANO-) CIE DANONE SA GERVAIS.
PA (RHOD) RHODIA CHIM.
XX
PI Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;
XX
DR WPI; 2002-017616/02.
XX
PT New nucleic acid fragments containing exopolysaccharide operon, useful
PT e.g. for increasing exopolysacchride synthesis in lactic acid bacteria.
XX
PS Claim 11; Page 48-52; 144pp; French.
XX
CC The present sequence is an eps operon from Streptococcus thermophilus.
CC Proteins encoded by the eps operon function in exopolysaccharide (EPS)
CC synthesis. The operon is useful for producing chimeric eps operons, for
CC optimising production of EPS in lactic acid bacteria. EPS impart texture,
CC mouth feel and rheological properties to fermented food products (e.g.
CC yoghurt). They function as thickeners, to provide free-flowing and creamy
CC texture, and may also have biological activities beneficial to health
XX
SQ Sequence 14142 BP; 4483 A; 2605 C; 2582 G; 4411 T; 0 U; 61 Other;

Query Match 5.4%; Score 54.8; DB 6; Length 14142;
Best Local Similarity 48.2%; Pred. No. 0.026;
Matches 189; Conservative 0; Mismatches 197; Indels 6; Gaps 1;

```
Qy         601 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 660

Db       13644 AATATTTTTGGTGTGCCAGAACGACCAAGATTTTTACCAGATTATTTTAATTACTTTCAA 13585

Qy         661 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 720

Db       13584 AAGGTGACACAAGAAGAATTATTACAGTTATATAATGATACAGCTATTTTTGTATGTGCT 13525

Qy         721 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 780

Db       13524 ACAATTGATGAAGGATTCGGACTCACAGGAGCGGAAAGCATGGCATGTGGATGCGCTCTT 13465

Qy         781 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 840

Db       13464 GCTTCGACTGCATATAGCGGAGTTTTTGAATACGCTGTGGACGGGGTAAATGCCTTGTTG 13405

Qy         841 CCTCTTTCGGAAGATGCTA------TTACTAAAGGAATTTCGAGGTTAATTAATGATTCT 894

Db       13404 TCTCCTGTTAGAGATAGTAGCCAACTTGCTAACAATATTATCAAGTTAATTGAGGATAAT 13345

Qy         895 GAACTTCGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAA 954

Db       13344 GAGCATCGTCAATTATTGGCAAAAAGAGCTAGTGATATGCTTATAGAGCGTTCATGGAAA 13285

Qy         955 AACGTGGTTAGTGAGATTGAAATGGTACTGAC 986

Db       13284 GCAACTATTAATAAGCTTGAAGAGGAATTGTC 13253
```

RESULT 14
ABA01443
ID ABA01443 standard; DNA; 20555 BP.
XX
AC ABA01443;
XX
DT 21-FEB-2002 (first entry)
XX
DE Streptococcus thermophilus eps5 operon #2.
XX
KW Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
KW ds.
XX
OS Streptococcus thermophilus.
XX
PN WO200179500-A2.
XX
PD 25-OCT-2001.
XX
PF 18-APR-2001; 2001WO-FR001199.
XX
PR 18-APR-2000; 2000FR-00004972.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
PA (DANO-) CIE DANONE SA GERVAIS.
PA (RHOD) RHODIA CHIM.
XX
PI Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;
XX
DR WPI; 2002-017616/02.
XX
PT New nucleic acid fragments containing exopolysaccharide operon, useful
PT e.g. for increasing exopolysacchride synthesis in lactic acid bacteria.
XX
PS Claim 9; Page 88-93; 144pp; French.

XX
CC The present sequence is an eps operon from Streptococcus thermophilus.
CC Proteins encoded by the eps operon function in exopolysaccharide (EPS)
CC synthesis. The operon is useful for producing chimeric eps operons, for
CC optimising production of EPS in lactic acid bacteria. EPS impart texture,
CC mouth feel and rheological properties to fermented food products (e.g.
CC yoghurt). They function as thickeners, to provide free-flowing and creamy
CC texture, and may also have biological activities beneficial to health
XX
SQ Sequence 20555 BP; 6613 A; 3185 C; 4041 G; 6716 T; 0 U; 0 Other;

Query Match 5.4%; Score 54.8; DB 6; Length 20555;
Best Local Similarity 48.2%; Pred. No. 0.028;
Matches 189; Conservative 0; Mismatches 197; Indels 6; Gaps 1;

Qy 601 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 660
Db 14310 AATATTTTGGTGTGCCAGAACGACCAAGATTTTTACCAGATTATTTAATTACTTTCAA 14369

Qy 661 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 720
Db 14370 AAGGTGACACAAGAAGAATTATTACAGTTATATAATGATACAGCTATTTTTGTATGTGCT 14429

Qy 721 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 780
Db 14430 ACAATTGATGAAGGATTCGGACTCACAGGAGCGGAAAGCATGGCATGTGGATGCGCTCTT 14489

Qy 781 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 840
Db 14490 GCTTCGACTGCATATAGCGGAGTTTTTGAATACGCTGTGGACGGGGTAAATGCCTTGTTG 14549

Qy 841 CCTCTTTCGGAAGATGCTA------TTACTAAAGGAATTTCGAGGTTAATTAATGATTCT 894
Db 14550 TCTCCTGTTAGAGATAGTAGCCAACTTGCTAACAATATTATCAAGTTAATTGAGGATAAT 14609

Qy 895 GAACTTCGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAA 954
Db 14610 GAGCATCGTCAATTATTGGCAAAAAGAGCTAGTGATATGCTTATAGAGCGTTCATGGAAA 14669

Qy 955 AACGTGGTTAGTGAGATTGAAATGGTACTGAC 986
Db 14670 GCAACTATTAATAAGCTTGAAGAGGAATTGTC 14701

RESULT 15
ADZ70999/c
ID ADZ70999 standard; DNA; 1554 BP.
XX
AC ADZ70999;
XX
DT 14-JUL-2005 (first entry)
XX
DE Human chromosome 2 contig DNA SEQ ID NO 123.
XX
KW ds; matrix attachment region; MAR; protein production.
XX
OS Homo sapiens.
XX
PN WO2005040377-A2.
XX
PD 06-MAY-2005.
XX
PF 22-OCT-2004; 2004WO-EP011974.
XX
PR 24-OCT-2003; 2003US-0513574P.
PR 06-FEB-2004; 2004EP-00002722.
XX
PA (SELE-) SELEXIS SA.
XX
PI Mermod N, Girod PA, Bucher P, Nguyen D, Calabrese D, Saugy D;
PI Puttini S;
XX
DR WPI; 2005-333507/34.

XX
PT New purified and isolated DNA sequence having protein production
PT increasing activity comprises a bent DNA element and a binding site for a
PT DNA binding protein, useful for increasing protein production activity in
PT eukaryotic host cell.
XX
PS Disclosure; SEQ ID NO 123; 282pp; English.
XX
CC The invention relates to a purified and isolated DNA sequence having
CC protein production increasing activity comprising at least one bent DNA
CC element, and at least one binding site for a DNA binding protein. The
CC purified and isolated DNA sequence comprising a first and a second
CC isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC MAR nucleotide sequence selected from a purified and isolated DNA
CC sequence above, a purified and isolated MAR DNA above, a purified and
CC isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC sequence complementary to it, its molecular chimera, or its combinations
CC and variants, is useful for increasing protein production activity in a
CC eukaryotic host cell. The present sequence represents a human chromosome
CC 2 contig DNA.
XX
SQ Sequence 1554 BP; 865 A; 14 C; 27 G; 648 T; 0 U; 0 Other;

Query Match 5.3%; Score 53.4; DB 14; Length 1554;
Best Local Similarity 43.1%; Pred. No. 0.034;
Matches 308; Conservative 0; Mismatches 406; Indels 1; Gaps 1;

Qy 1 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 60
Db 1031 ATATTTAATATATAGTTATATTTAATATATATTTAATATATAGTTATATTTAATATATAT 972

Qy 61 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 120
Db 971 TTAATATATAGTTATATTTAATATATATTTAATATATAGTTATATTTAATATATATTTAA 912

Qy 121 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 180
Db 911 TATATAGTTATATTTAATATATATTTAATATATAGTTATATTTAATATATATTTTAATA 852

Qy 181 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 240
Db 851 TATATATTTATATTTAATATATATTTTTAATATATATATTTTATATATATTTTTAATATA 792

Qy 241 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 300
Db 791 TATATTTTATATATATTTTTAATATATATATTTTATATATATTTTTAATATATATATTTT 732

Qy 301 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 360
Db 731 ATATATATTTATATTATATATATTTTTAATATATATATTTTATATATATTTATATTATAT 672

Qy 361 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 420
Db 671 ATATTTTTAATATATATATTTTATATATATTCATATTTAATATATATTTTATATATATTT 612

Qy 421 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 480
Db 611 ATATTTAATATACTATATTTAATACAT-ATATTTATATTCTATATATATAATACATATAT 553

Qy 481 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 540
Db 552 TTATATTTAATATATATTTAATACATATATTCATATTTAATATATATTTAATACATATAT 493

Qy 541 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 600
Db 492 TTATATTTAATATATATATTTAATACATATATTTATATTTTATATATATTTATATTTAAT 433

Qy 601 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 660
Db 432 ATATATATTTAATATATATATATTTTATATATATTTAATATATATTTATATTTTATATAT 373

Qy 661 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTT 715

Db 372 CTTTTATATATTTAACATATATATTTTATATATATTTAATATATATTTATATTTT 318

Search completed: May 11, 2006, 09:42:47
Job time : 464.665 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 09:43:19 ; Search time 3149.94 Seconds
(without alignments)
15016.694 Million cell updates/sec

Title: US-09-423-093B-3_COPY_6313_7323
Perfect score: 1011
Sequence: 1 atgaataataaaaaggtttt.....catgtgatggaaataaatga 1011

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	85.8	8.5	551	10	AG947336	AG947336 Drosophil
	2	76.4	7.6	260	10	AG942630	AG942630 Drosophil
c	3	63.4	6.3	1101	10	CNS0039G	AL063921 Drosophil
c	4	62.2	6.2	1101	10	CNS00EVL	AL069706 Drosophil
c	5	61.8	6.1	1101	10	CNS00FVE	AL071298 Drosophil
	6	60	5.9	928	10	CNS00DKY	AL071865 Drosophil
	7	59.2	5.9	1092	10	CNS020K7	AL175696 Tetraodon
	8	58	5.7	1027	10	CNS02T50	AL212733 Tetraodon
	9	57.6	5.7	1101	10	CNS004ZW	AL055440 Drosophil
c	10	57.4	5.7	1161	11	CNS073Y8	AL428118 clone BA0
c	11	57	5.6	1146	10	CNS021G2	AL176843 Tetraodon
c	12	56.6	5.6	1101	10	CNS003BD	AL064091 Drosophil
	13	56.4	5.6	987	10	CNS014PQ	AL104456 Drosophil
c	14	56.2	5.6	830	9	BZ556422	BZ556422 pacs1-60_
c	15	56.2	5.6	1101	10	CNS00EO7	AL069440 Drosophil
	16	56	5.5	1101	10	CNS00EVL	AL069706 Drosophil
	17	55.8	5.5	807	10	CL489403	SAIL_523_
	18	54.8	5.4	1101	10	CNS0039G	AL063921 Drosophil
c	19	54.6	5.4	1043	10	CNS0145P	AL103735 Drosophil
	20	54.6	5.4	1101	10	CNS003BD	AL064091 Drosophil
c	21	54.6	5.4	1101	10	CNS00BO1	AL057419 Drosophil
	22	54.2	5.4	693	9	BH992718	BH992718 oeh28g06.
	23	54.2	5.4	959	10	CNS017OR	AL108309 Drosophil
	24	54.2	5.4	1086	10	CNS00YXK	AL096962 Drosophil
	25	54	5.3	905	10	CNS00KHX	AL077798 Drosophil
	26	53.8	5.3	499	1	AU037114	AU037114 AU037114
	27	53.8	5.3	570	1	AU033518	AU033518 AU033518
c	28	53.8	5.3	622	3	BJ435345	BJ435345 BJ435345
c	29	53.8	5.3	651	3	BJ399848	BJ399848 BJ399848
c	30	53.8	5.3	662	3	BJ435757	BJ435757 BJ435757
	31	53.8	5.3	1101	10	CNS00EO7	AL069440 Drosophil
	32	53.6	5.3	720	9	BH983915	BH983915 odg44g08.
	33	53.6	5.3	1101	10	CNS00BO1	AL057419 Drosophil
c	34	53.6	5.3	1101	10	CNS00EPO	AL069493 Drosophil
	35	53.6	5.3	1101	10	CNS00Z2U	AL097152 Drosophil
c	36	53.4	5.3	796	7	CK474754	CK474754 AGENCOURT
	37	53.2	5.3	945	11	CNS04D0K	AL285149 Tetraodon
	38	53	5.2	968	9	BZ562608	BZ562608 pacs2-164
	39	53	5.2	1201	10	CNS016A0	AL106482 Drosophil
	40	52.8	5.2	950	8	DN564456	DN564456 88035777
c	41	52.8	5.2	1027	10	CNS02T50	AL212733 Tetraodon
	42	52.8	5.2	1101	10	CNS00LT2	AL078714 Drosophil
	43	52.6	5.2	771	1	AL521957	AL521957 AL521957
	44	52.6	5.2	826	10	AG579614	AG579614 Mus muscu
	45	52.4	5.2	1101	10	CNS0106X	AL098595 Drosophil

ALIGNMENTS

RESULT 1
AG947336
LOCUS AG947336 551 bp DNA linear GSS 01-FEB-2005
DEFINITION Drosophila ananassae DNA, clone: DNB1-024M11.R.fa, genomic survey sequence.
ACCESSION AG947336
VERSION AG947336.1 GI:58468949
KEYWORDS GSS.
SOURCE Drosophila ananassae
ORGANISM Drosophila ananassae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1
AUTHORS Hattori,M., Toyoda,A., Murakami,K., Kuroki,Y., Fujiyama,A., Toshio,T.K. and Sakaki,Y.
TITLE BAC end sequences of Library DNB1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 551)
AUTHORS Hattori,M.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2005) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the BAC library DNB1
For BAC library availability, please contact Masa-Toshi Yamamoto (yamamoto@kit.jp).
Submitted (30-11-2004) by Masahira Hattori, RIKEN, Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.jp,Tel:81-45-503-9111, Fax:81-45-503-9170)
This work was done in collaboration with Yamamoto, M-T. Drosophila Genetic Resource Center
Saga Ippongi-cho, Ukyo-ku, Kyoto 616-8354, Japan
Tel: 81-75-873-2660 FAX: 81-75-861-0881
PRIMERS
Sequencing : R
LIBRARY
Vector : pKS150
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES Location/Qualifiers

```
     source          1. .551
                     /organism="Drosophila ananassae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7217"
                     /clone="DNB1-024M11.R.fa"
                     /clone_lib="DNB1 Drosophila BAC library"
ORIGIN

  Query Match             8.5%;  Score 85.8;  DB 10;  Length 551;
  Best Local Similarity   51.4%;  Pred. No. 6.1e-09;
  Matches  198;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy        539 CTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTT 598
              |||| ||   ||| | |     |     ||| | | || || |||| |   |
Db         18 CTGCATTCAATAAATTACCTGAAAAGTATGAAATGCATATGATTGTTGCTGGGAGCACGC  77

Qy        599 GTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTG 658
              | | | |   |     || | | | |     | || |    || |  ||| | |
Db         78 ATCCTAACTATCAGCCCTTAGAATTTGATAAAAACCTCGGTAATAAAGTCAGTTTCGTCG 137

Qy        659 GGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCC 718
              ||   || |||||  | ||  |       | |||      ||||  | ||  ||||||
Db        138 GGCGAGTTTCTGACGAGGAGCTTGTGGATCTGTATCGCAATGCGTTCTGCTTTCTTTTCC 197

Qy        719 CTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTG 778
              | ||  |||||||||| ||||| || ||     | || |     |||||||||  ||| |
Db        198 CATCCCTATATGAAGGCTTTGGTTTGCCACCGCTAGAAGCTCAAGCCTGTGGTTGTCCCG 257

Qy        779 TATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCG 838
              | |  | |||     |  |  | |||| || ||    ||| ||    || ||||||||||
Db        258 TGATTGCATCTAATGCCACTGCGATGCCTGAAGTTCTTGGCGAAAGTGCAATTCTTGTCG 317

Qy        839 ACCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAAC 898
              | ||  |    || ||||  | ||| |||  ||  || | |  ||  |||  |  |||
Db        318 ATCCATTAAACGATGATGCATTAACTGAAGCTATGGAGAAGCTGGTTCTTGACCCAAAAC 377

Qy        899 TTCGTAAGCATTTAATCCAAAAGGG 923
              |  |  ||  | | |||  |   ||
Db        378 TCAGAGAGTCTCTTATCGCAGCAGG 402


RESULT 2
AG942630
LOCUS       AG942630               260 bp    DNA     linear   GSS 01-FEB-2005
DEFINITION  Drosophila ananassae DNA, clone: DNB1-016G04.R.fa, genomic survey
            sequence.
ACCESSION   AG942630
VERSION     AG942630.1  GI:58464049
KEYWORDS    GSS.
SOURCE      Drosophila ananassae
  ORGANISM  Drosophila ananassae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Hattori,M., Toyoda,A., Murakami,K., Kuroki,Y., Fujiyama,A.,
            Toshio,T.K. and Sakaki,Y.
  TITLE     BAC end sequences of Library DNB1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 260)
  AUTHORS   Hattori,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2005) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the BAC library DNB1
            For BAC library availability, please contact Masa-Toshi Yamamoto
            (yamamoto@kit.jp).
            Submitted (30-11-2004) by  Masahira Hattori,
            RIKEN, Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp,Tel:81-45-503-9111,
            Fax:81-45-503-9170)
            This work was done in collaboration with Yamamoto, M-T. Drosophila
            Genetic Resource Center
            Saga Ippongi-cho, Ukyo-ku, Kyoto 616-8354, Japan
            Tel: 81-75-873-2660 FAX: 81-75-861-0881
            PRIMERS
                 Sequencing : R
            LIBRARY
                 Vector     : pKS150
                 R.Site 1   : SacI
                 R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1. .260
                     /organism="Drosophila ananassae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7217"
                     /clone="DNB1-016G04.R.fa"
                     /clone_lib="DNB1 Drosophila BAC library"
ORIGIN

  Query Match             7.6%;  Score 76.4;  DB 10;  Length 260;
  Best Local Similarity   56.3%;  Pred. No. 8e-07;
  Matches  143;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        646 GTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTA 705
              || | ||| | |||   || |||||  | ||  |       | |||      ||||
Db          6 GTCAGTTTCGTCGGGCGAGTTTCTGACGAGGAGCTTGTGGATCTGTATCGCAATGCGTTC  65

Qy        706 GGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCC 765
              | ||  ||||||| ||  |||||||||| ||||| || ||      | || |      |||
Db         66 TGCTTTCTTTTCCCATCCCTATATGAAGGCTTTGGTTTGCCACCGCTAGAAGCTCAAGCC 125

Qy        766 TGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCA 825
              ||||||  |  |||  |  | |||     |  |  |  |||| || ||   || ||
Db        126 TGTGGTTGTTTTGTGATTGCATCTAATGCCACTGCGATGCCTGAAGTTCTTGTCGAAAGT 185

Qy        826 GCGATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATT 885
              || |||||||||| ||  |    || |||||  | ||| |||  ||   || | |  ||
Db        186 GCAATTCTTGTCGATCCATTAAACGATGATGCATTAACTGAAGCTATGGAGAAGCTGGTT 245

Qy        886 AATGATTCTGAACT 899
              |||  |  ||||
Db        246 CTTGACCCAAAACT 259


RESULT 3
CNS0039G/c
LOCUS       CNS0039G               1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
```

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 6.3%; Score 63.4; DB 10; Length 1101;
Best Local Similarity 16.9%; Pred. No. 0.0009;
Matches 115; Conservative 282; Mismatches 281; Indels 1; Gaps 1;

```
Qy          3 GAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGTTT 62
              ::: ::|:::: :::::: :   ::::: ::::: :|:::| ::| | ::     :::
Db       1098 RRWGDDTWDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWW 1039

Qy         63 TACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAATG 122
              ::   :: :::::: :| | :  ||: : :||::  :| :::|    :::|  :|: | :
Db       1038 WWWATWDTWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGR 979

Qy        123 TGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACTGA 182
               :       : :::    ::: |::::|  | : | ::::::: ::   ::|| ::  ::
Db        978 RDGGRKRKDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADD 919

Qy        183 TGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATAAC 242
              ::  |::::: :    :: |:  ::  :    ::::: : | |:| :|    ::::: :
Db        918 DDGAGDKDDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWA 859

Qy        243 AATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTATT 302
              :|::  ::|  ::::  :  :  |:::::|: : |:|     : :   :: ||::  : :
Db        858 DADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRK 799

Qy        303 TTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCGAA 362
              :: ::  :: :: | :       ||:: ::: :: :|:|::||::|::: :::  :|
Db        798 RADDKRDAADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAG 739

Qy        363 TTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACGGT 422
              |   : :: ::::::::: :::: |: :: |:::: ::     :|     | :: : ::
Db        738 TAGRKWRRTWKRRWKRRDTRWDD-ADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDR 680

Qy        423 ATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTATAA 482
              ||::: :    :|  :  | |:|:::::  :::| ::::      |:::: : :::|
Db        679 ATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAK 620

Qy        483 ATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCTGC 542
              ::::|    ::::  :   :|   ::::|   |::|:::  : |: ::  ||   | |
Db        619 WDWKTRADRWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTTAT 560

Qy        543 CTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGTAA 602
               ||:   ||:  : |:||| ||   |   |:    |  ||||:  |  :::   :| |:|
Db        559 WTTWAARAAWWAWWAWATTTATWTTTWTTTWTTTTTTTTTTTTWTTTAAWWWAAWWTATWA 500

Qy        603 TGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTT 662
                  : ||||| :|||  |       |   :||| :  |  : |  : || ::    ||
Db        499 AWTAAAWAAAAAAAAWAATTTTTTTTTTTTTWAWTAAAWWTAWTWTTTTWTTTTWWAATTTT 440

Qy        663 CGTGTCTGAAAAAGATTTA 681
              : : | :||    |||:
Db        439 TTWTWTTWWAATTATTTTW 421
```

RESULT 4
CNS00EVL/c

LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"

ORIGIN

Query Match 6.2%; Score 62.2; DB 10; Length 1101;
Best Local Similarity 35.0%; Pred. No. 0.0017;
Matches 169; Conservative 86; Mismatches 224; Indels 4; Gaps 2;

```
Qy        158 ATATTTTTCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTT 217
              |:| :| | :   ::|  :|| : || :  |||           :|:::|     |||
Db        984 AWAAWTATAWAWTWWAATWAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTT 925

Qy        218 TTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATA 277
              ||| ||     :  |: :|    | ::|: |  |: ::||: :| |   ||:  || ||
Db        924 TTTTTTTWTTTWTATWTWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATT 865

Qy        278 ATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAG 337
              : |  | : ||||| : :  ||::| |||:||::||| :|: :       :| |: |
Db        864 WATATTAWATTTTTTWTWTATTWWTATATWATWWTAT-WTWTWTTTWATWTATWTATTTA 806

Qy        338 CATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTG 397
               :|:|:  :|||||:     | |  ::|:: | :: | ||: |     |:         :
Db        805 TWTWTWTTWTATTWTATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWA 746

Qy        398 TAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATG 457
              | :    |      : |     :   |:|||:||    :|| |  | | ||| |    ||
Db        745 TTWTTWTTTTATTWTTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTWTATT 686

Qy        458 TAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATA 517
              |:|||  ||| |||| :     :|: |||||  ||  |      ||  |  |||| |  ||
Db        685 TWAAATTATTTAATTAWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTA 626

Qy        518 AGAATGAGAAGTGTGTTATAT---CTGCCTTTGCCAAAGCAGATATTGATCCATCAATAA 574
                ||  |  | | | ||||||    :  |||   :|   | |||::  :  |   :| |
```

```
Db        625 TWAAAAAATATTTTTTTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTA 566

Qy        575 AACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTT 634
              |  |  |::||     || :  || |  | :| :| : :|| :|||:|:: :: |   |:
Db        565 ATWTATTWWTTTAAATTATWTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATW 506

Qy        635 TAA 637
              :::
Db        505 WWW 503


RESULT 5
CNS00FVE/c
LOCUS       CNS00FVE                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR32P18 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071298
VERSION     AL071298.1  GI:4951138
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR32P18"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             6.1%;  Score 61.8;  DB 10;  Length 1101;
  Best Local Similarity   30.0%;  Pred. No. 0.0021;
  Matches  183;  Conservative  148;  Mismatches  273;  Indels    7;  Gaps    3;

Qy          2 TGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGTT 61
              :|: : :|: : ::|: |:  :::|: ||  :   | || |  :   |: :|| :    |
Db       1070 KGWTWTDTDGRGRRGWATWWTWDKAWTTTTTKGTTTTAGTGKWTWTGRAWTTTKGWAAT 1011

Qy         62 TTACTGATGAAATTTCTAAACTACTATG--TGATATATCTAAGGAGGAACTATATAGAAA 119
              : :  ::| ::||:  |||  |:  : |  | | |:|| | |  |::   ||::|:| |:
Db       1010 DGDTADRTADRATWAATAAKGTWAGRRGGATAAAAWATATKATWAKRGKTTADDTDGKAW 951

Qy        120 ATGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAAC 179
              ::::  ::    : : |    |:: : ||:|  |:  : : ||    :| ||:   :||
Db        950 WWDWATWWWGARDGKAGWATAAWDGRWTTWATAAKGTWTWTTTGGGGKGGGAWGATWAAA 891

Qy        180 TGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAAT 239
```

```
              :|:|:    :        : : | ||:   :      :  :|:: |  :|:| | |:||
Db        890 KGRTDAWWAKGAGGGGAAWWKATTTAWAGTWTTGRAADGAWGWDWGTKDADGRTDADAAA 831

Qy        240 AACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTT 299
              :: :| : : |:: :::|| :  | :|||:||  |||   :    :  : ::  :| :||
Db        830 WDADAAWAWAGWWDWWRATRWGATTRATTDAATAGATTTGWWKGTKTAWAWWTKKAWKTT 771

Qy        300 ATTTTATAATTTTATAATAAAGCGCGGTTGTAGA----AAAGCATATAAAATATTTACAG 355
              : :  :| ||| |  |||::|:     :|:|| |    | || | :||| || | |: |:
Db        770 WGWAAWTTATTGTGAAATRRARAAAATWTRTAAAAGAGATAGAARWTAAGATKTATWTAD 711

Qy        356 TTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAG 415
              | | ::|||::|   |  |::: |::| |  ::  | : :|:  |:   :||  :| |
Db        710 TATWKWATTKWTTTTATAAWWKTAKWGGA-AWKWATTWTDTRATAWAWKWAAATWATTTT 652

Qy        416 TCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAG 475
              : |    |:|| :   || || :||| |: :  :|    |::|::    ||| || |||
Db        651 KAATAWAADATTWAAAGGGATWWAGTKTWGWGKWTWKDAAWKTWWTTGGATTAAAATTAT 592

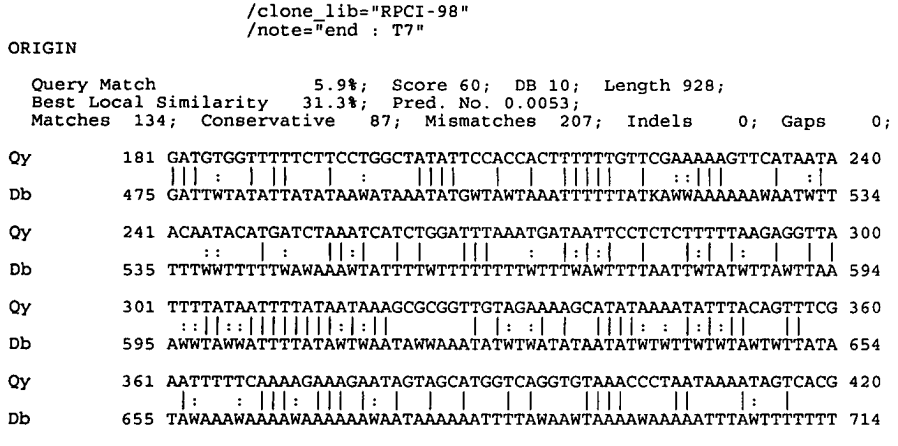
Qy        476 GCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTA 535
              | | || :|  |:: |   |  : :|    :|:: : | : |   |   :||  |:| :
Db        591 GGTTTATWTTRTKKWTTAATWAWADATTTGKAWWTDWTTWAATTTTTTTAWAGATTKTGR 532

Qy        536 TATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATC 595
              | | |:  : ::      | :| ||  :  || :  | :    |  ||   | |  |
Db        531 TTTTTVTAWGWRGATGTATATWTTTTATWARATARNGATHTTGTGGNTTGTTTATTTTTN 472

Qy        596 CTTGTAATGAT 606
              ||||  |  |
Db        471 TTTGTTTTNTT 461

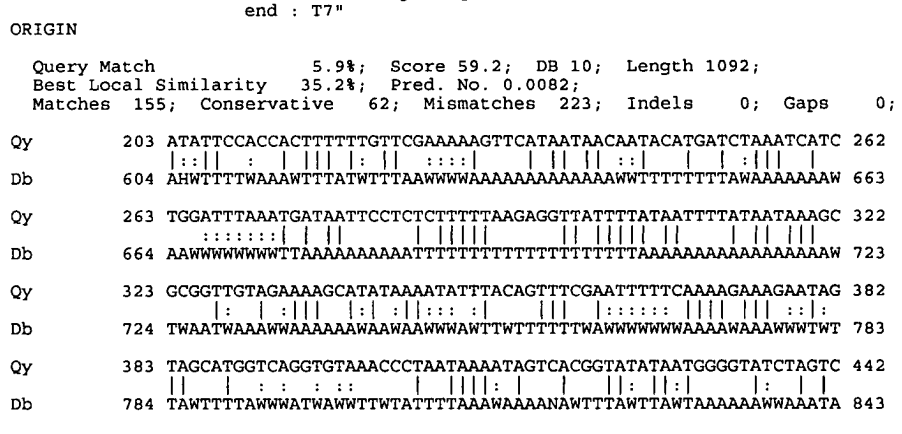

RESULT 6
CNS00DKY
LOCUS       CNS00DKY                928 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR27A24"
```

/clone_lib="RPCI-98"
/note="end : T7"
ORIGIN



Query Match 5.9%; Score 60; DB 10; Length 928;
Best Local Similarity 31.3%; Pred. No. 0.0053;
Matches 134; Conservative 87; Mismatches 207; Indels 0; Gaps 0;

RESULT 7
CNS020K7
LOCUS CNS020K7 1092 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 222L11 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL175696
VERSION AL175696.1 GI:7813753
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
PUBMED 10899143
REFERENCE 3 (bases 1 to 1092)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .1092
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="222L11"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG222CF06LP1
end : T7"
ORIGIN



Query Match 5.9%; Score 59.2; DB 10; Length 1092;
Best Local Similarity 35.2%; Pred. No. 0.0082;
Matches 155; Conservative 62; Mismatches 223; Indels 0; Gaps 0;

RESULT 8
CNS02T50
LOCUS CNS02T50 1027 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 163M16 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL212733
VERSION AL212733.1 GI:7871552
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1

```
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1027)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1027
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="163M16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG163BG08LP1
                     end : T7"
ORIGIN

  Query Match             5.7%;  Score 58;  DB 10;  Length 1027;
  Best Local Similarity   40.4%;  Pred. No. 0.015;
  Matches  193;  Conservative   34;  Mismatches  250;  Indels    1;  Gaps    1;

Qy        276 TAATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAA 335
Db        544 TATTTTTTTWTTTAATAATATAATTTTTNTAAATTWATAAATAAAAWAATTATATATATA 603

Qy        336 AGCATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGG 395
Db        604 AWAATTTTTTATATATTTTATTTTATATTATNNATAATTATTTTTTTTTTWWWWWWWWWW 663

Qy        396 TGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGA 455
Db        664 AAAAAAAAAAAAAAAAWAAAAARTRRATTGTTTTWTTWTWWWWWWWATWTWWTAWGTATT 723

Qy        456 TGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTG-TAGGAAACAGAAAAACTC 514
Db        724 TNTAAAAATAAAAAAATAAAATAATTAAATTGTTAAATNTAATANTAGATWTWTAATATA 783

Qy        515 ATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAA 574
Db        784 TTTATTAWATWRWGWATGTNTNATAAAATTWTAWAATNATTTWGATAGATTAAAAAKTRT 843

Qy        575 AACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTT 634
Db        844 TWTTTTDTTTTTGTTTKTTGGTGTGTGTGTGTTTGAAAATTTTTTTAATTTTTTWTGATT 903

Qy        635 TAAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATA 694
Db        904 TATTTGATTTTSTTTATTTTTTATAAATTATATATTTAATAWAATATTTNTTTTTATATA 963

Qy        695 AGGGCTCGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGT 752
Db        964 AKTTTTTAAGTGTTTTTTTTTTGATATRTATTTTTTTAATTTTATTTATTTTTGTAAT 1021
```

```
RESULT 9
CNS004ZW
LOCUS       CNS004ZW                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR11E08 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL055440
VERSION     AL055440.1  GI:4932241
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR11E08"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             5.7%;  Score 57.6;  DB 10;  Length 1101;
  Best Local Similarity   39.6%;  Pred. No. 0.019;
  Matches  116;  Conservative   38;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        213 ACTTTTTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAA 272
Db        716 ATTTNTAAAKTAAAAAAWATWTATATWTMWAATWWTTTAWATWTAAAATTWAWWTAWATA 775

Qy        273 TGATAATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAG 332
Db        776 TATATWWNTATWTWTTWWWWTTATNTTATTTTTATATTTTATATATAASAAAVAATATWT 835

Qy        333 AAAAGCATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTC 392
Db        836 TAAATMAAATANTATANTTATATTATTTAAATTTATTCATWAATATAATTTTTAATATTA 895

Qy        393 AGGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGC 452
Db        896 WATMAAATAAWTAAAATWTTTAWTATATWTTTWTAWTTTTKTTATWTTAWDATWWTATTW 955

Qy        453 CGATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACA 505
Db        956 ATATWTATATWWATAAAAWWTAAWTWWAWATTTATATATWTAWTAAAAAAWTA 1008


RESULT 10
CNS073Y8/c
LOCUS       CNS073Y8                1161 bp    DNA     linear   GSS 07-JUL-2001
```

```
DEFINITION  clone BA0AB023G11 of library BA0AB from strain CLIB 210 of
            Kluyveromyces lactis, genomic survey sequence.
ACCESSION   AL428118
VERSION     AL428118.1  GI:12211312
KEYWORDS    GSS.
SOURCE      Kluyveromyces lactis
  ORGANISM  Kluyveromyces lactis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   2  (bases 1 to 1161)
  AUTHORS   Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F.,
            Duchateau-Nguyen,G., Lemaire,M., Marmeisse,R., Montrocher,R.,
            Robert,C., Termier,M., Wincker,P. and Wesolowski-Louvel,M.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 11.
            Kluyveromyces lactis
  JOURNAL   FEBS Lett. 487 (1), 66-70 (2000)
   PUBMED   11152886
REFERENCE   3  (bases 1 to 1161)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1161
                     /organism="Kluyveromyces lactis"
                     /mol_type="genomic DNA"
                     /strain="CLIB 210"
                     /variety="lactis"
                     /db_xref="taxon:28985"
                     /clone="BA0AB023G11"
                     /clone_lib="BA0AB"
ORIGIN

  Query Match             5.7%;  Score 57.4;  DB 11;  Length 1161;
  Best Local Similarity   36.8%;  Pred. No. 0.021;
  Matches  214;  Conservative   61;  Mismatches  300;  Indels    6;  Gaps    2;

Qy        213 ACTTTTTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAA 272
              | : :||  ||  || || ||  :||||  | |      : ||| | || :  ||| |||
Db        834 AAWAHTTATTTWWAATAAYTTTTHAATATAATTTATAATYATAACTTATTYWAATTAAAA 775

Qy        273 TGATAATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAG 332
              |  |  |   | | : ||| |      |   | || ||| ||| |||       ||
Db        774 TNTTYTTAATTTTWTAWATAATAWTAATTAAAAAAAATTTTTAAAAAATATAATTTAANA 715

Qy        333 AAAAGCATA--TAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGG 390
              || |  |||  | |||||||||   |||  |||  ||  |||  |    |  |   ||
Db        714 AATATTATATTTTAAATATTTAATTTTTTTAATAATTTTAAAATATTTTTTTTYCTATTA 655

Qy        391 TCAGGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAAT 450
                |  | |||||  ||| || : | | |     ||:||       : || | |||| :||

Db        654 AAAAATATAAACAATAAAAATWAAATTAT----TAWAACAANAACBATATTATATTAWAT 599

Qy        451 GCCGATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAA 510
                | :|  |  | |  ||||      |    | |::: : : :: |  || :| |||
Db        598 TTTATTAKANTATTAATTTATTTTAAAAAATTTTTTKKKKKKKTBSCACCAACYAAAAAN 539

Qy        511 ACTCATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCA 570
              |: |             | | |: | | |   |       |  | |||||  :  |  |
Db        538 TTTMYTYCCCCCCCCCTTTTTTTWTTTTTTTTNTNANNANNATAANATATTTTWTAAAAA 479

Qy        571 ATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACAT 630
              : |   |  ::  :     | ||  |  |||| :|: | |     ::  |::| || ||
Db        478 TWTATTTTTTWWAAWTAACYTKATTTTATTAATTWTWAAAATTYYAKWTAAWDAAAAAAT 419

Qy        631 GGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTA 690
              ||  | ||   :||   :| ::::::|:  : | :  |   : |||   |    |
Db        418 AAAATATTTTAAAAAKTATBTDTDKKKKKKTWWADTTTBTYABSKKTTTTNNTANNNATG 359

Qy        691 TATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTA 750
              ||||    : ::   :||     |:|  |:| |  || : :| |       :|   | |
Db        358 TATATCTASSSSSSKKGANATTGTWTNTCBTGTAAATNBNWATGAANNTATWTTTATTTT 299

Qy        751 GTGGAGGGCATGGCCTGTGGTATTCCTGTATTAACTTCTCT 791
              |        :  : | :::| | ::|:  |      | |||
Db        298 WTAATTTTTTYTTYTTTKKBTNTSBBTSSCTNNNNNTNTCT 258


RESULT 11
CNS021G2/c
LOCUS       CNS021G2                1146 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            225004 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL176843
VERSION     AL176843.1  GI:7814900
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1146)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1146
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
```

/clone="225O04"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG225BH02LP1
end : T7"

ORIGIN

Query Match 5.6%; Score 57; DB 10; Length 1146;
Best Local Similarity 30.8%; Pred. No. 0.026;
Matches 178; Conservative 107; Mismatches 291; Indels 1; Gaps 1;

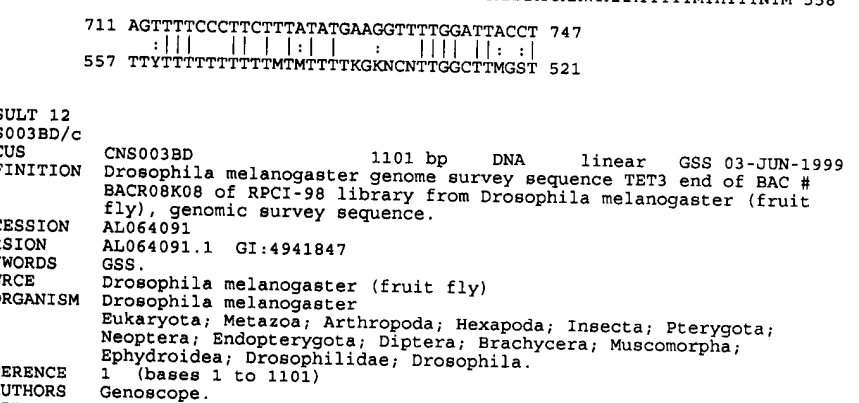


RESULT 12
CNS003BD/c
LOCUS CNS003BD 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL064091
VERSION AL064091.1 GI:4941847
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K08"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 5.6%; Score 56.6; DB 10; Length 1101;
Best Local Similarity 37.3%; Pred. No. 0.032;
Matches 240; Conservative 66; Mismatches 331; Indels 7; Gaps 2;



RESULT 13

CNS014PQ
LOCUS CNS014PQ 987 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN12P22 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL104456
VERSION AL104456.1 GI:5616067
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 987)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .987
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN12P22"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : SP6"
ORIGIN

Query Match 5.6%; Score 56.4; DB 10; Length 987;
Best Local Similarity 27.5%; Pred. No. 0.036;
Matches 138; Conservative 123; Mismatches 239; Indels 1; Gaps 1;

```
Qy          46 GGGGGGATTGGACGTTTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAG 105
               |||:::  : :: |    :   ||   :    :|:|  | :::: ||:: |:  ||| |
Db         461 GGGRKRTRKAKRKGSGGGRGATATANGRGGGAWAWAAAAADRDRRGAWDAAWKAAAGAAA 520

Qy         106 GAACTATATAGAAAATGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTT 165
                 | | ||     |||: :| | :          :| ::   :: : |  | | :  :
Db         521 ATTCACTYTAWKGCGTGTRTKTGTGMTAGWTDATRGWTGRKGGRDKDWTTAGAAAAWAAD 580

Qy         166 CTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCG 225
                ::::::::   :  :::::::|:: ||  | : ::: |: | || |          :
Db         581 RAARDRRRRRRBKKBNKKKKKKKKTKKTTTTTTTKAYMWCTBKCCCCCCCCCCCMMMAMM 640

Qy         226 AAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCT 285
               |:||:  ::|::| :| ||| |  ||  |||  |:    |  |||| |: ||:  : |
Db         641 AMAAMAAWMAWWACWAAAAT-TAGAATAAAAAAAWAWAAAAAAAAAAATAAWTATWWAWNT 699

Qy         286 CTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAA 345
                :|:||:: :  || |||| |    : | |::|||      :|  : | ::    |  ||
Db         700 NWTWTTWWWWTWTTTTTTTTTTTTAAWAAAAWWAAAWAAAWTWTTAWWATWWTTTAAAGAA 759

Qy         346 ATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCT 405
               |: | |  | ||:  | ||||:   :|  ||  | :  |  :   :      :::|
Db         760 AWTTATTAAATTWAAATTTTTWWTTWATTAATTAAWTAATTWATTWATTWATWWWAATTA 819

Qy         406 AATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCA 465
               |  : ::||   :   : | |  |     |||  | | | :::|:     : | : ::  |
Db         820 ATAWTWWTANAKWTTWWTTTTTTTTTTWWTATAAATTWTTWWWAWAAATAWRGRWTWWAWA 879

Qy         466 TTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAG 525
                | || ::|    | |:|:::::::  |::: ::   :  | | :  ||   | || ||: |
Db         880 ATWAAAWWATWWAAAAWAWWWWDRAAAWDWWWWWAWTAAAWWTTAAAAAAAAAAAAWAAA 939
```

```
Qy         526 AAGTGTGTTATATCTGCCTTT 546
               || : : :::::: :   |||
Db         940 AAAWWWWWWWWWWWTTTTTT 960
```

RESULT 14
BZ556422/c
LOCUS BZ556422 830 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs1-60_5670.x1 pacs1-60 Pseudomonas aeruginosa genomic clone pacs1-60_5670, genomic survey sequence.
ACCESSION BZ556422
VERSION BZ556422.1 GI:27167231
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 830)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .830
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacs1-60_5670"
/clone_lib="pacs1-60"
/note="clinical isolate 1-60 Whole genomic shotgun library."
ORIGIN

Query Match 5.6%; Score 56.2; DB 9; Length 830;
Best Local Similarity 56.9%; Pred. No. 0.039;
Matches 103; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy         646 GTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTA 705
               ||   | |  | || | | || | ||  |     |  |   |||      |||| |
Db         285 GTGCGGCTGCTCGGCTACCTGCCGGACGAGCGGGTCGCCGAGTTGATGTCGGGCGCTCGG 226

Qy         706 GGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCC 765
               |   | || ||||| ||  | || ||||| || ||  | |||||  |||| | |||||||
Db         225 GCGCTGGTCTTCCCCTCGATCTACGAAGGCTTCGGCCTGCCTGTGCTGGAAGCCATGGCC 166

Qy         766 TGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCA 825
                | || |  || ||||      | | |   ||  || |||| |||||||| || || ||
Db         165 AGCGGCACGCCGGTATTGCTCACGCGACTCTCCGCAATGCCGGAGGTGGCCGGCGAGGCG 106

Qy         826 G 826
               |
Db         105 G 105
```

RESULT 15
CNS00EO7/c
LOCUS CNS00EO7 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069440
VERSION AL069440.1 GI:4949583

```
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29P01"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.6%;  Score 56.2;  DB 10;  Length 1101;
  Best Local Similarity   36.6%;  Pred. No. 0.04;
  Matches  219;  Conservative   73;  Mismatches  299;  Indels    7;  Gaps    2;

Qy        155 TCAATATTTTTCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCAC 214
              | |:|||||||  |   :|| |       || |    |   :: :    ::||| |
Db       1010 TTAWTATTTTTTTTTTTMAATATYCWATATTTTMAATACAWWAYATTTWWTATACATAATW 951

Qy        215 TTTTTTGTTCGAAAAAGTTCATAATAACAATACATGA---TCTAAATCATCTGGATTTAA 271
              :|||||  |  | ||  |: | |  || |:||  : |   | :|||:||| :   || |:
Db        950 WTTTTTTATATACAATTTWAAAATAAAAAAWTAACWWAATTTAWAAAWCATTWTTTTTAAW 891

Qy        272 ATGATAATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTA 331
              :| ||  :    |  |:|| :: |    | |:|  ||:|  |:|| ||:        |
Db        890 WTAATTTWATTACAWTWTTAWWTAAAAAAATWTTAAAWTAAAWAAAAARWATTAWAATTT 831

Qy        332 GAAAAGCATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGT 391
              | :  ||:||||::  | :    ||: :| |:||  |:         |  ||  |:  |
Db        830 AYATWATATWTAAAWWTATAWATTATTMWWAATWTTATAWNATTTTTTTTTWTAWAAWTWT 771

Qy        392 CAGGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATG 451
              :  : : ||   |||: :: ||   |    | ||       ||| |: | | :  |:|
Db        770 TWATWAWTAATTTTAAWTWWTTAAATAAAAAAAATTTATTTTTATTTWTTATTWAAAWTT 711

Qy        452 CCGATGTAAAACCATTG----AATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGA 507
              | |:||   ::|     |:|| |   : ||||:|::|  : |   ||   :
Db        710 TTWTTTTWAATTWWYTTTAATAWTTAAAWTWTTAAAWAWWTAAWTTAAATAAATWTATTT 651

Qy        508 AAAACTCATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCA 567
              |||| : : |: |   |    : : |||||: |   |||  |:    | ||||  |   |
Db        650 AAAATWTWAAWTATAAATTTAWADWTTATAWWTTTTTTTWWCWTTWAATATATAAAATWA 591

Qy        568 TCAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAA 627
              |||  |||  |  ||||   |  :|         ::  | |  | ||:  | || |  :|
Db        590 AAAATTAAATTWTTTTTATATWWWATAAAAAMATWWWTTATTTATAAWWAATTATTTAWA 531

Qy        628 CATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCG 687
              || |||   |  |  | |    |   :|    |  | |      |   ||:|    |
Db        530 WTTGATTTTTATTTATTTWTTTTTTAAAWTTTTATTATATWATTTATTAATWTTAWTTAT 471

Qy        688 TTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTAC 745
              :| | |      || ||   |:|| || |||| ::|||:|      ||||   ||:|
Db        470 YTTTTTTTTATATCTTTCMAAWTATTTMTCCCCYYTTTWTTMATGTTTTTTTTTTTHC 413

Search completed: May 12, 2006, 03:29:46
Job time : 3154.94 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 19:15:19 ; Search time 138.274 Seconds
(without alignments)
12996.800 Million cell updates/sec

Title: US-09-423-093B-3_COPY_6313_7323
Perfect score: 1011
Sequence: 1 atgaataataaaaaggtttt.....catgtgatggaaataaatga 1011

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	270.6	26.8	1029	3	US-09-489-039A-987	Sequence 987, App
	2	74	7.3	1161	3	US-09-489-039A-1153	Sequence 1153, Ap
	3	61.6	6.1	1182	3	US-09-252-991A-2472	Sequence 2472, Ap
	4	58.8	5.8	1664976	3	US-08-916-421B-1	Sequence 1, Appli
	5	58.8	5.8	1664976	3	US-09-692-570-1	Sequence 1, Appli
	6	53.6	5.3	30820	3	US-09-949-016-17145	Sequence 17145, A
	7	53.4	5.3	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	8	50	4.9	25992	3	US-09-949-016-17308	Sequence 17308, A
c	9	49.8	4.9	55886	3	US-09-949-016-15129	Sequence 15129, A
c	10	49.6	4.9	1664976	3	US-08-916-421B-1	Sequence 1, Appli
c	11	49.6	4.9	1664976	3	US-09-692-570-1	Sequence 1, Appli
c	12	49.4	4.9	50000	3	US-09-662-254B-25	Sequence 25, Appl
c	13	49.2	4.9	1141	3	US-09-806-708B-22	Sequence 22, Appl
	14	49	4.8	640681	3	US-09-790-988-1	Sequence 1, Appli
	15	47.6	4.7	832	3	US-09-621-976-2813	Sequence 2813, Ap
	16	47	4.6	1410	3	US-09-662-254B-33	Sequence 33, Appl
	17	47	4.6	50000	3	US-09-662-254B-24	Sequence 24, Appl
	18	46.8	4.6	1164	3	US-09-601-198-65	Sequence 65, Appl
c	19	46.6	4.6	601	3	US-09-949-016-156535	Sequence 156535,
c	20	46.6	4.6	32392	3	US-09-662-254B-27	Sequence 27, Appl
	21	46.6	4.6	50000	3	US-09-662-254B-23	Sequence 23, Appl
	22	46.6	4.6	231129	3	US-09-949-016-16110	Sequence 16110, A
	23	46.6	4.6	266293	3	US-09-949-016-11934	Sequence 11934, A
	24	46.2	4.6	660	2	US-07-991-867B-32	Sequence 32, Appl
	25	46.2	4.6	660	2	US-08-107-755A-32	Sequence 32, Appl
	26	46.2	4.6	660	2	US-08-544-332-32	Sequence 32, Appl
	27	46.2	4.6	660	3	US-09-370-861A-32	Sequence 32, Appl
	28	46.2	4.6	1511	2	US-07-991-867B-8	Sequence 8, Appli
	29	46.2	4.6	1511	2	US-08-107-755A-8	Sequence 8, Appli
	30	46.2	4.6	1511	2	US-08-544-332-8	Sequence 8, Appli
	31	46.2	4.6	1511	3	US-09-370-861A-8	Sequence 8, Appli
	32	46.2	4.6	4810	3	US-08-852-629-11	Sequence 11, Appl
	33	46.2	4.6	4838	3	US-08-852-629-15	Sequence 15, Appl
c	34	46.2	4.6	50000	3	US-09-662-254B-23	Sequence 23, Appl
c	35	46	4.5	640681	3	US-09-790-988-1	Sequence 1, Appli
	36	45.8	4.5	187169	3	US-09-949-016-12776	Sequence 12776, A
	37	45.8	4.5	191569	3	US-09-949-016-15940	Sequence 15940, A
	38	44.4	4.4	50000	3	US-09-662-254B-25	Sequence 25, Appl
c	39	44	4.4	700	3	US-09-735-271-1039	Sequence 1039, Ap
	40	43.8	4.3	1032	3	US-09-902-540-6426	Sequence 6426, Ap
	41	43.8	4.3	2632	3	US-09-902-540-440	Sequence 440, App
c	42	43.8	4.3	50000	3	US-09-662-254B-26	Sequence 26, Appl
c	43	43.6	4.3	50000	3	US-09-662-254B-24	Sequence 24, Appl
	44	43.6	4.3	119153	3	US-09-949-016-12378	Sequence 12378, A
	45	43.4	4.3	1077	3	US-09-328-352-1480	Sequence 1480, Ap

ALIGNMENTS

RESULT 1
US-09-489-039A-987
; Sequence 987, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 987
; LENGTH: 1029
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-987

Query Match 26.8%; Score 270.6; DB 3; Length 1029;
Best Local Similarity 56.2%; Pred. No. 1e-55;
Matches 554; Conservative 0; Mismatches 419; Indels 12; Gaps 2;

```
Qy          1 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 60
              |||   || || ||  | ||  | |||| | |||||     | |    || || ||  |
Db         43 ATGGGAAAAAAGAAAATATTATTTGATACTCGTTGGATAGGTGAGCATGGTATAGGTAGA 102

Qy         61 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 120
              ||| || | ||| ||     |    |  ||    ||||     |     |||||     | |||
Db        103 TTTGCTAAGGAAGTTAGAGACTCAAAATTAACTATACAGGACTTAAAACTACCATGGAAA 162

Qy        121 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 180
                | |   |||||| |   || ||   ||  |  |  || |  || |       ||||
Db        163 CCTACAAATCCGCTTGATACAATATTATTGACTATTTATTTAATTTTT------AAAAAA 216

Qy        181 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 240
              ||| |   ||||   ||||| || |   | ||  ||||||     |||   | |||||
Db        217 GATATCTATTTTTCACCTGGATACAACGCGCCTTTTTTTTTCCTTAAACGATCTATAATC 276

Qy        241 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 300
              || || |||||||| |||||| | ||| |   | ||||    ||  || |||| ||  |
Db        277 ACGATTCATGATCTTAATCATATAGATATTCCTTATAACAGTTCCTTTCTTAAAAGAATT 336

Qy        301 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 360
              | |||||| ||| |  | ||  | | ||| |||||||| |      ||| | || || ||
Db        337 TATTATAAGTTTGTGCTTAAAAGAGCTTGCAGAAAAGCGTTATGTATACTAACCGTATCT 396
```

```
Qy        361 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 420
              | ||||| || | || ||    | ||||||  |||  ||||  |  ||  ||  |   |
Db        397 GAATTTTCTAAGAAGAGGATTTGTGAATGGTCGGGTCGAAACGCGGATGAAGTTTTTGTA 456

Qy        421 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 480
              |||   |||||  || || |     |||   |     ||     | ||    ||   ||
Db        457 GTAGGAAATGGAGTGTCCAAAGAGTTTCTTATAAATCACCCAACACCCAAGAAAG----- 511

Qy        481 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 540
              || |  |   ||   || |||| ||||  ||||| |||||    |   ||    ||
Db        512 -AACAAATATTAATGGTTGGTAACAGGAAAAAACATAAAAATGAATTGAATGCACTAATA 570

Qy        541 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 600
              ||  |||    ||  || ||||     ||||  | || || ||||||||  |||  |  |
Db        571 GCTTTTTATAAGTCAAATATACCCTCATCGGTTAAGCTAGTTTTTACAGGTTCTATTACC 630

Qy        601 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 660
              ||||| |||||  | | | ||     |   |  |  | || ||  | |||    ||  ||||
Db        631 AATGAATTAGATATAATTATCGATCGTAACAGGCTTAGGGATAGGGTATCTTTTCTTGGT 690

Qy        661 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 720
              || ||  | |    |  | |    ||||||||  |  || |||   ||  | || |||
Db        691 ACAGTATCAAATACTAAACTTGCCGAGTTATATAGAGAATCCTTATTTTTGCTATTTCCT 750

Qy        721 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 780
              || ||||||||||||||  ||  | ||  |||||||   |||| || || | ||| ||
Db        751 TCGTTATATGAAGGTTTCGGTCTCCCGATAGTGGAGTCAATGGCTTGCGGAACTCCAGTT 810

Qy        781 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 840
              |||  |||   |   |||| || || || || |||||||  || || ||| ||||  ||
Db        811 ATAATCTCTAATAGCACATCTTTACCTGAAGTTGCTGGAGGGGCTGCTATTATTGTAAAC 870

Qy        841 CCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTT 900
              ||   |   | ||| | ||| |   |    ||||   |  |||  | |||||  | ||||||
Db        871 CCAGATGATGTAGAGGATATAAAAGATAAAATTGAAAAATTATATCATGATGCAGAACTT 930

Qy        901 CGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTG 960
              |  ||  |  | ||| |||||||  |||||         | || |  |  ||| |||
Db        931 CAAAATAAACTTATCGAAAAGGGTCTTTTAAATATTGAAAGATATACCTGGGAAATGACT 990

Qy        961 GTTAGTGAGATTGAAATGGTACTGA 985
              | | |  ||||||   || |||
Db        991 CGCAATAAATTTGAAACAATAATGA 1015
```

RESULT 2
US-09-489-039A-1153
; Sequence 1153, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1153
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1153

Query Match 7.3%; Score 74; DB 3; Length 1161;
Best Local Similarity 52.2%; Pred. No. 2.3e-08;
Matches 164; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```
Qy        611 AAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTG 670
              ||  || | ||||| ||    ||     | ||||  | |  ||| | |  |  ||    |
Db        767 AATCACCATTAATAAAATTACGTAATGATAAACGCTTTATATTCATCGATTATGTCGAAG 826

Qy        671 AAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTTATATG 730
              ||      |||   | | | ||||||     | ||  | ||  |||   || ||||||  |||
Db        827 GAAGTACATTAGATTGGCTTTATAAGAATGCATTTTGTTTTATTTATCCCTCTTTAAATG 886

Qy        731 AAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTAACTTCTC 790
              |||| | ||| |  ||      |  ||| ||||   |  ||||  || ||| |  | |||
Db        887 AAGGATATGGTTATCCGCCGTTACAGGCCATGAGGTACGGTACACCCGTAATTGCATCTT 946

Qy        791 TAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGACCCTCTTTCGG 850
              || |||    | ||||| ||   || ||||   |    |  | | || ||
Db        947 GCACATCAAGTATTCCAGAAGTATGTGCAGATAATGTACTATATTTTGATCCAAGGAATA 1006

Qy        851 AAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTTCGTAAGCATT 910
              |  |  ||    ||  |||||  |   |||  ||  |||  |||||||  ||| |
Db       1007 TAAGCGAAATCGAAAATCGAATTCTGCAATTATCTACAGATGGTGAACTTTATAATAAAC 1066

Qy        911 TAATCCAAAAGGGG 924
              ||      ||||||
Db       1067 TAGCTTTGAAGGGG 1080
```

RESULT 3
US-09-252-991A-2472
; Sequence 2472, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2472
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2472

Query Match 6.1%; Score 61.6; DB 3; Length 1182;
Best Local Similarity 50.0%; Pred. No. 2.2e-05;
Matches 154; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

```
Qy        646 GTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTA 705
              ||   | |  | || | | || | ||  |     |   |   |||       |||| |
Db        823 GTGCGGCTGCTCGGCTACCTGCCGGACGAGCGGGTCGCCGAGTTGATGTCGGGCGCTCGG 882

Qy        706 GGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCC 765
              |   | || ||||| ||  | || ||||| || ||  | |||||  |||| | |||||||
Db        883 GCGCTGGTCTTCCCCTCGATCTACGAAGGCTTCGGCCTGCCTGTGCTGGAAGCCATGGCC 942

Qy        766 TGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCA 825
              | || |  || |||||     | | |   ||  || |||| |||||||| || || ||
Db        943 AGCGGCACGCCAGTATTGCTCACGCGACTCTCCGCAATGCCGGAGGTGGCCGGCGAGGCG 1002

Qy        826 GCGATTCTTGTCGACCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATT 885
              |  |     | | |||||    |    |   |||   ||  |      |     |  | |
Db       1003 GGCAGCTATATTGACCCGGATGATGCCAATGGGCTATCTGACATGCTGGTCCGCATGACC 1062

Qy        886 AATGATTCTGAACTTCGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTC 945
              | ||| | |      |  ||    |      ||   || ||    |||||  ||  |||
Db       1063 GAGGATACCGTCTACTGGGAGTGCTGCCGGGAAGCCGGCCTGCGACGGGCGGAGCAGTTT 1122

Qy        946 AATTGGCA 953
              |||  |
Db       1123 TCCTGGAA 1130
```

```
RESULT 4
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1130881)..(1130881)
```

```
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349491)..(1349491)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

  Query Match             5.8%;  Score 58.8;  DB 3;  Length 1664976;
  Best Local Similarity   48.7%;  Pred. No. 0.00059;
  Matches  224;  Conservative    0;  Mismatches  227;  Indels    9;  Gaps    2;

Qy        546 TGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGA 605
              ||| | |  |||||| |||        | |||||| |      ||| ||    |  ||  |
Db     355500 TGCAATAAAAGATATAGATTTTAATTTTAAACTCATAGGGGATGGGAAGTTATACAAAAA 355559

Qy        606 TTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGT 665
                ||||  ||  |  |  |  || ||  ||||||    | | |  |  |||| ||
Db     355560 AATAGAGAACTTTGTTGTTAAAAATAATTTAAG---CCATATTGAACTCTTAGGAAGAAA 355616

Qy        666 GTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTT 725
                |    | ||  || | || ||    | |   |          |||||  | |||||
Db     355617 AAGTTTTGATGAAGTAGCTTCATTTATGAGGAAGTGTAGTTTTTTAGTAGTTCCTTCAAG 355676

Qy        726 ATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTAAC 785
              |  |||||||||||||| |    |  || || || ||||| ||   ||  |||||| || |
Db     355677 AAGTGAAGGTTTTGGAATGGTGGCTGTTGAAGGAATGGCTTGCTCTAAGCCTGTAATAGC 355736

Qy        786 TTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGA------TTCTTGTCGA 839
                |      |      |||   ||| | | |   |||| |    |      |  | | ||
Db     355737 CACAAGGGTTGGGGGGTTGGGGGAGATTGTTATTGATGGATATAACGGACTATTGGCTGA 355796

Qy        840 CCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACT 899
                |    | ||| | | |   ||  ||||  |  |||||| ||| ||   |||||
Db     355797 GAAAAATAACCCAAATGATTTAAAAGAAAAAATTCTGGAGTTAATAAATAATGAAGAACT 355856

Qy        900 TCGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGT 959
              | ||   |||      |||| ||          |||| |  ||   |||| ||||
Db     355857 AAGAAAAACTTTGGGGGAAAATGGAAAAGAATTTTCAAAAAAATTTTCTTGGGAAAAATG 355916

Qy        960 GGTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGAT 999
              || |  |   ||  ||  ||      ||||  | | |||
Db     355917 TGTAATGGGTGTTAGAAAAGTGTATGAAGAGCTAAGCGAT 355956


RESULT 5
US-09-692-570-1
; Sequence 1, Application US/09692570
```

```
; Patent No. 6797466
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococc
; Patent No. 6797466
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275C1
;  CURRENT APPLICATION NUMBER: US/09/692,570
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  PRIOR APPLICATION NUMBER: US 08/916,421
;  PRIOR FILING DATE: 1997-08-22
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
```

; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g

Query Match 5.8%; Score 58.8; DB 3; Length 1664976;
Best Local Similarity 48.7%; Pred. No. 0.00059;
Matches 224; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

```
Qy        546 TGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGA 605
              ||| | |  |||||| |||        | |||||| |      ||| ||    |  ||  |
Db     355500 TGCAATAAAAGATATAGATTTTAATTTTAAACTCATAGGGGATGGGAAGTTATACAAAAA 355559

Qy        606 TTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGT 665
              |||| ||  |  |  |  || ||  ||||||    | | |  |  |||| ||
Db     355560 AATAGAGAACTTTGTTGTTAAAAATAATTTAAG---CCATATTGAACTCTTAGGAAGAAA 355616

Qy        666 GTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTT 725
              |    | ||  || | || ||     | |   |         |||||  | |||||
Db     355617 AAGTTTTGATGAAGTAGCTTCATTTATGAGGAAGTGTAGTTTTTTAGTAGTTCCTTCAAG 355676

Qy        726 ATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTAAC 785
              |  |||||||||||||| |    |  || || || ||||| ||   ||  |||||| || |
Db     355677 AAGTGAAGGTTTTGGAATGGTGGCTGTTGAAGGAATGGCTTGCTCTAAGCCTGTAATAGC 355736
```

```
Qy        786 TTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGA------TTCTTGTCGA 839
Db     355737 CACAAGGGTTGGGGGGTTGGGGGAGATTGTTATTGATGGATATAACGGACTATTGGCTGA 355796

Qy        840 CCCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACT 899
Db     355797 GAAAAATAACCCAAATGATTTAAAAGAAAAAATTCTGGAGTTAATAAATAATGAAGAACT 355856

Qy        900 TCGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGT 959
Db     355857 AAGAAAAACTTTGGGGGAAAATGGAAAAGAATTTTCAAAAAAATTTTCTTGGGAAAAATG 355916

Qy        960 GGTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGAT 999
Db     355917 TGTAATGGGTGTTAGAAAAGTGTATGAAGAGCTAAGCGAT 355956
```

RESULT 6
US-09-949-016-17145
; Sequence 17145, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17145
; LENGTH: 30820
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17145

Query Match 5.3%; Score 53.6; DB 3; Length 30820;
Best Local Similarity 43.6%; Pred. No. 0.0041;
Matches 239; Conservative 0; Mismatches 309; Indels 0; Gaps 0;

```
Qy        141 ATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGG 200
Db       6046 ATTATATACTATATATAATATATAAATATATATTATATACTATATATAATATATAAATAT 6105

Qy        201 CTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCA 260
Db       6106 ATATTATATACTATATATAATATATAAATATATATTATATACTATATATAATATATAAAT 6165

Qy        261 TCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAA 320
Db       6166 ATATATTATATACTATATATAATATATAATATATACTATATTATATATAGTATATTATAT 6225

Qy        321 GCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAAT 380
Db       6226 ACTATATTATATAGTATAATATAGAATATATTATATATTGTATATAAGATAATATATTAT 6285

Qy        381 AGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAG 440
Db       6286 ATATTGTATATAAGATATAATATATTATATATTGTATATAAGATATAATATATTATATAT 6345

Qy        441 TCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGG 500
Db       6346 TGTATATAAGATAATATATTATATATTGTATATAAGATATAATATATTATATATTGTATA 6405

Qy        501 AAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATAT 560
```

```
Db       6406 TAAGATATAATATATTATATATTGTATATAAGATATAATATATTATATATTGTATATAAG 6465

Qy        561 TGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAAT 620
Db       6466 ATATAATATATTATATATTGTATATAAGATATAATATATTATATATTGTATATAAGATAT 6525

Qy        621 AATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTT 680
Db       6526 AATATATTATATATTGTATATAAGATATAATATATTATATATTGTATATAAGATATATTA 6585

Qy        681 ACCATCGT 688
Db       6586 TATATTGT 6593
```

RESULT 7
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 5.3%; Score 53.4; DB 3; Length 1141;
Best Local Similarity 12.7%; Pred. No. 0.0021;
Matches 91; Conservative 238; Mismatches 386; Indels 3; Gaps 1;

```
Qy         66 TGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAATGTGC 125
Db        315 KVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNWTYKKARHBARWDWVWHSAW 374

Qy        126 TTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACTGATGT 185
Db        375 KKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYNNNNNGGRT 434

Qy        186 GGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATAACAAT 245
Db        435 YYGWTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKTYWNNCYWKSMTNGKSHR 494

Qy        246 ACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTATTTTA 305
Db        495 BAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBVCSKWWNNYAAWYTKSSWNYTSR 554

Qy        306 TAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCGAATTT 365
Db        555 YYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWSHT---WBHBRAGAAHYW 611

Qy        366 TTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACGGTATA 425
Db        612 MBMMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNNNNNATCARDDYYAASRWYAM 671

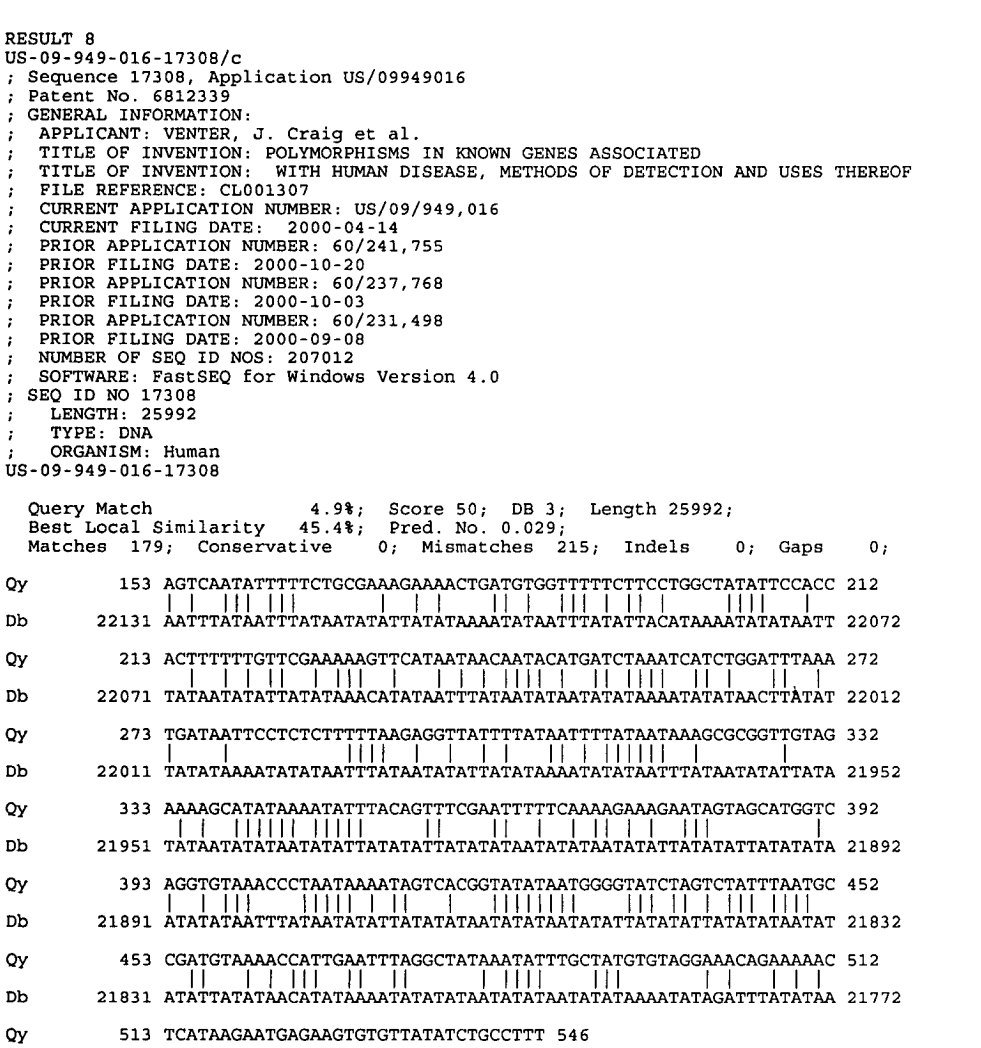
Qy        426 TAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTATAAATA 485
Db        672 ANAKWYYYKBAANNAYYTHANNWWGCWNNATDTRRTMWKNNNNNNAGTWKNNNNNNAKNA 731

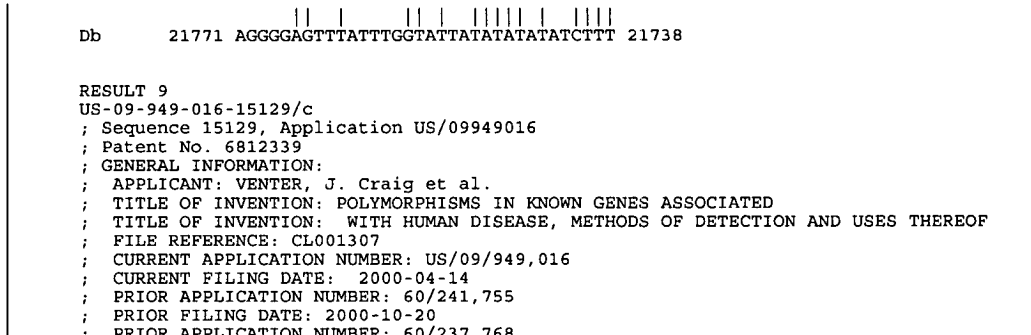
Qy        486 TTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCTGCCTT 545
```



RESULT 8
US-09-949-016-17308/c
; Sequence 17308, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17308
; LENGTH: 25992
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17308

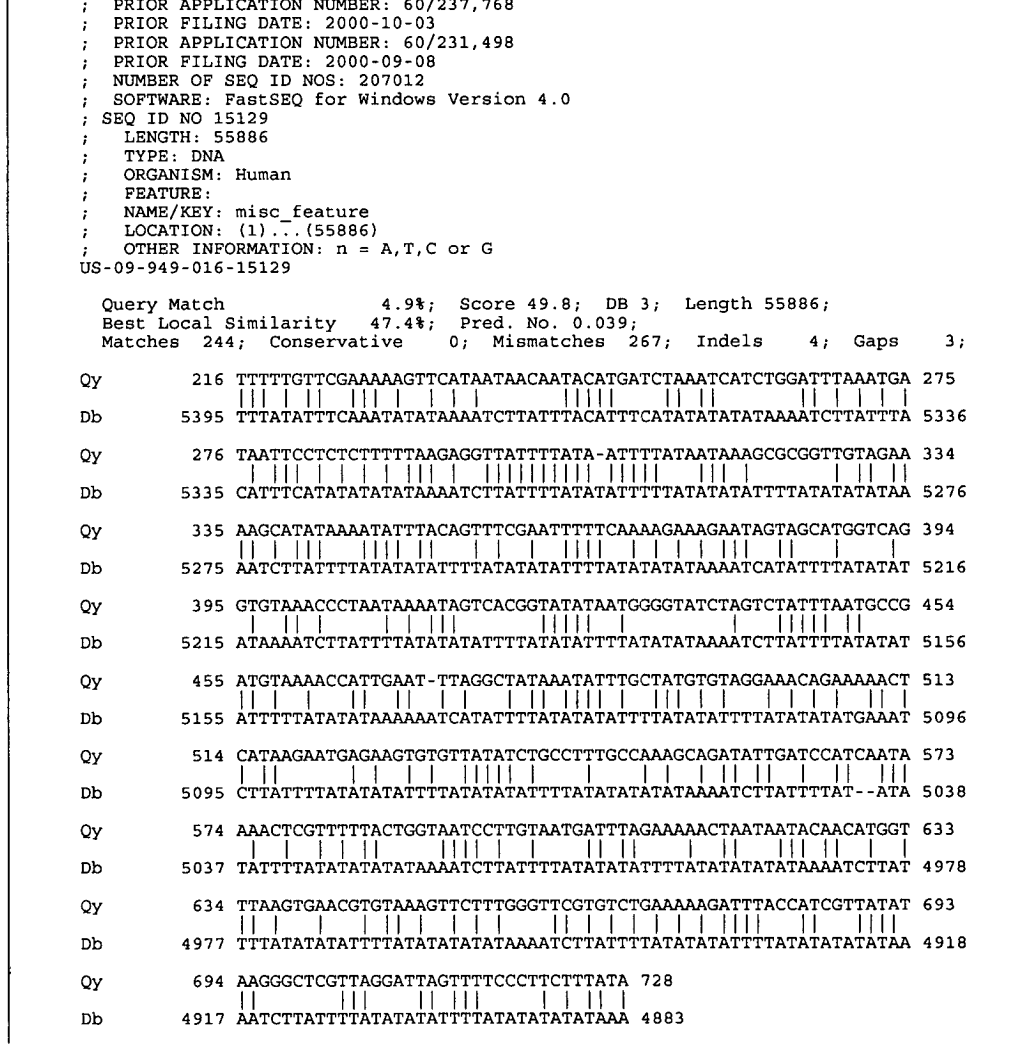
Query Match 4.9%; Score 50; DB 3; Length 25992;
Best Local Similarity 45.4%; Pred. No. 0.029;
Matches 179; Conservative 0; Mismatches 215; Indels 0; Gaps 0;





RESULT 9
US-09-949-016-15129/c
; Sequence 15129, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15129
; LENGTH: 55886
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(55886)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15129

Query Match 4.9%; Score 49.8; DB 3; Length 55886;
Best Local Similarity 47.4%; Pred. No. 0.039;
Matches 244; Conservative 0; Mismatches 267; Indels 4; Gaps 3;



```
RESULT 10
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (682442)..(682442)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (713652)..(713652)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (741684)..(741684)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779455)..(779455)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (779676)..(779676)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (855539)..(855539)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (871619)..(871619)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1084830)..(1084830)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1096846)..(1096846)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1119881)..(1119881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (1130881)..(1130881)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1310988)..(1310988)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1349491)..(1349491)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1470091)..(1470091)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1569020)..(1569020)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1602912)..(1602912)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1603734)..(1603734)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1637998)..(1637998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (1664854)..(1664855)
;   OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

  Query Match             4.9%;  Score 49.6;  DB 3;  Length 1664976;
  Best Local Similarity   50.0%;  Pred. No. 0.098;
  Matches  155;  Conservative    0;  Mismatches  149;  Indels    6;  Gaps    1;

Qy        479 ATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATAT 538
              | |||    |  ||| ||| ||||    || |  ||| |||       ||| | | |||
Db    1555002 AAAAAATGATTTTATTTGTTGGAAGATTAACATATCAAAAGGGAATAGAGTATTTAATAA 1554943

Qy        539 CTGCCTTTGCCAAA------GCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTA 592
              ||  |  | |||       | |      | || ||  | ||    | |    | ||||
Db    1554942 GAGCAATGCCAAAAATTCTTGAAAGACATAATGCAAAATTAGTTATTGCAGGTTCTGGAG 1554883

Qy        593 ATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGT 652
              ||    |  || |||| ||| |  |        || | |||| | ||  |    |||  ||
Db    1554882 ATATGAGGGATTATTTGGAAGATTTGTGCTATCAGCTTGGTGTTAGGCATAAGGTAGTGT 1554823

Qy        653 TCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAG 712
              | || || || ||   || | |    |||  |    | |||||      ||   |   | |
Db    1554822 TTTTAGGATTTGTTAATGGAGATACGTTAAAAAAACTCTATAAATCTGCGGATGTTGTTG 1554763

Qy        713 TTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTA 772
              |  | || ||  | |||||     |||||  ||  ||    | ||||  |||||   ||| |
Db    1554762 TAATTCCATCAGTTTATGAGCCATTTGGTATAGTTGCTTTAGAGGCAATGGCTGCTGGCA 1554703

Qy        773 TTCCTGTATT 782
              || ||| |
Db    1554702 CACCAGTAGT 1554693


RESULT 11
US-09-692-570-1/c
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6797466
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275C1
;  CURRENT APPLICATION NUMBER: US/09/692,570
;  CURRENT FILING DATE:  2003-01-14
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  PRIOR APPLICATION NUMBER: US 08/916,421
;  PRIOR FILING DATE: 1997-08-22
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
```

; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g

Query Match 4.9%; Score 49.6; DB 3; Length 1664976;
Best Local Similarity 50.0%; Pred. No. 0.098;
Matches 155; Conservative 0; Mismatches 149; Indels 6; Gaps 1;

```
Qy        479 ATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATAT 538
              | |||     |  ||| ||| ||||    || |  ||| |||       ||| | | |||
Db    1555002 AAAAAATGATTTTATTTGTTGGAAGATTAACATATCAAAAGGGAATAGAGTATTTAATAA 1554943

Qy        539 CTGCCTTTGCCAAA------GCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTA 592
                ||  | | |||      | |      | || ||  | ||    | |    | ||||
Db    1554942 GAGCAATGCCAAAAATTCTTGAAAGACATAATGCAAAATTAGTTATTGCAGGTTCTGGAG 1554883

Qy        593 ATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGT 652
              ||      |  || |||| ||| |  |       || | |||| | ||  |   |||  ||
Db    1554882 ATATGAGGGATTATTTGGAAGATTTGTGCTATCAGCTTGGTGTTAGGCATAAGGTAGTGT 1554823

Qy        653 TCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAG 712
              | || || || ||   || | |     |||  |    | |||||     ||   |   | |
Db    1554822 TTTTAGGATTTGTTAATGGAGATACGTTAAAAAAACTCTATAAATCTGCGGATGTTGTTG 1554763

Qy        713 TTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTA 772
              |  | ||  ||  | |||||    |||||  ||  ||   | ||||  |||||   ||| |
Db    1554762 TAATTCCATCAGTTTATGAGCCATTTGGTATAGTTGCTTTAGAGGCAATGGCTGCTGGCA 1554703

Qy        773 TTCCTGTATT 782
                ||  ||| |
```

```
Db    1554702 CACCAGTAGT 1554693


RESULT 12
US-09-662-254B-25/c
; Sequence 25, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  APPLICANT:  Bawden, Alison Louise
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XC1
;  CURRENT APPLICATION NUMBER: US/09/662,254B
;  CURRENT FILING DATE:  2000-09-14
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-25

  Query Match             4.9%;  Score 49.4;  DB 3;  Length 50000;
  Best Local Similarity   47.1%;  Pred. No. 0.047;
  Matches  152;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy        117 AAAATGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAA 176
              || || |||| | |          |||| |||||    | |||| | |       ||  |
Db      42125 AATATTTGCTCATACAGACATAGAATTATGTTTATATTTAAATAATATAGAATATAAAGA 42066

Qy        177 AACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCAT 236
              | |      |  ||  | |||    || |  | | |      ||  || || || ||
Db      42065 AGATATAAATTTAATTAATTCTGATAGTAATATATTATTGAAACATTATAATAATTTAAT 42006

Qy        237 AATAACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAG 296
              ||||  |    || ||  | || |  |   |  | |   |||   || |  | ||| |
Db      42005 AATAGAATGTGATAATGAATATAAAATAATTAATAGTATAAATATTTCACAATATAAAAA 41946

Qy        297 GTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGT 356
                ||  | |||||  |||||| |        || |      |||| | || |    |
Db      41945 CATAACTAATAATAGAATAATATATTATAAACATAAATGTTTTAATAATAAATCTTGTTT 41886

Qy        357 TTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGT 416
              |    || ||||| ||||   |   ||  ||  | |  | |||     || ||||||  |
Db      41885 GTTTGTTTATTCAATAGAATATAATATATTATATTTATTTTTAATAAAAAAAAAAATATT 41826

Qy        417 CACGGTATATAATGGGGTATCTA 439
              |  ||  |  |    |||| ||
Db      41825 AAGTGTTAAGGAATATGTATATA 41803


RESULT 13
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             4.9%;  Score 49.2;  DB 3;  Length 1141;
  Best Local Similarity   16.7%;  Pred. No. 0.021;
  Matches   77;  Conservative  127;  Mismatches  258;  Indels    0;  Gaps    0;

Qy        232 TTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTT 291
              ||  :  : |         |:  :: :  || ::   | ::   |::  |:|:::::| |
Db       1129 TTKTKTYKKANNNNNNNNGMGKDWNRMDATKWSATGTAWWTNHAKRGATMCWYWYWTGTN 1070

Qy        292 AAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTT 351
              ::   :|:|  |: |:     :  | : ::       |::: |: |    :    ::  |
Db       1069 RRWCMRTYAMRTWYTRSNANWSCATKBMWWTMKWYATKYRTAWYAMWCAWRNNNMWCATN 1010

Qy        352 ACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAA 411
              :|:      | :: | ::|| |||::| :    ::  |  || :   :| : |: :::|
Db       1009 GYAKSCATNNAMWYATTRWAAYAAAKWARWAGNNMRMYGAAAGNKWGCMAAMATMGBWWA 950

Qy        412 ATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAAT 471
              :|||::       :| : ::            |::|| :       ::| : : :|  |:
Db        949 DTAGKMCNNNNNNWTTDVRRMAMKAKNNNNNNAYWTACYNRAATNNKMATHWMKWTHGAH 890

Qy        472 TTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGT 531
              :::  ::|:   :|::       :|:   :: |:::: | : :|           |   |
Db        889 SKRRTRHHTRTCRRTKYNNNNNNARTVYWYHHAARRWMNAWWTRTNNNNNNNNNNACRNT 830

Qy        532 GTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGT 591
              :|::  : :: |           |: :  |  |::    |          |      :
Db        829 RTWWABWKHSWCNNNNNNNNNNNNNNNTWCHYTTANABBCYRANNNNAAARMARTCNNYMH 770

Qy        592 AATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAG 651
              ||   |: :: :: :     :::: |: |  |: :   |  |       :||
Db        769 AAVTTTHTDWCYKTWMNTWYWDMMTTMBTTTTRNMTTSTNMTNNNNNNMWACTNNNNNNM 710

Qy        652 TTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATAT 693
              :: ::       |  :   :|::   || : ' :::  | |::
Db        709 WKAYYAHATNNWGCWWNNTDARRTNNTTVMRRRWMTNTKTRW 668


RESULT 14
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             4.8%;  Score 49;  DB 3;  Length 640681;
  Best Local Similarity   45.8%;  Pred. No. 0.11;
  Matches  169;  Conservative    0;  Mismatches  200;  Indels    0;  Gaps    0;

Qy        267 TTTAAATGATAATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGG 326
```

```
Db     140155 TATATATGACAATTTTTTTATTTTTATGTCTATTTAGTATAAATAAGAAAAATCTGCTAT 140214

Qy        327 TTGTAGAAAAGCATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGC 386

Db     140215 TTTAAAAAAAAATTTTTTTTAATTTACTTAGTAAAAGCAATCGATTTAACCGTATTTTAGA 140274

Qy        387 ATGGTCAGGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATT 446

Db     140275 ATCAGAATGATATATTTTAACTTTCTTAAAAAAATTATATATGGGATTTTCAAAACTTTT 140334

Qy        447 TAATGCCGATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAG 506

Db     140335 TATTTTTATTAAAATATCATTATATTTTTTTTCTAAAAATAAATTTTTTGTATAATTATC 140394

Qy        507 AAAAACTCATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCC 566

Db     140395 AAATTCTTCTATATTATTAAATAATATGATTTCTTCTTTTTCAATCATTAGTTTTTTATT 140454

Qy        567 ATCAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACA 626

Db     140455 AATATCTCCATTAATTTTTTCTTTTTCTTTTGCTAAAATGTTAGATATTCTTTTAATAGA 140514

Qy        627 ACATGGTTT 635

Db     140515 CAATATTAT 140523


RESULT 15
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

  Query Match             4.7%;  Score 47.6;  DB 3;  Length 832;
  Best Local Similarity   14.8%;  Pred. No. 0.048;
  Matches   60;  Conservative  172;  Mismatches  171;  Indels    3;  Gaps    1;

Qy         11 AAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGTTTTACTGATG 70

Db          2 RWYWWKYTTWYAKCWTKWKWSWSYWMYWKWYYMKTYWRWRRKKKKAWWKYWKTWTWWYWR 61

Qy         71 AAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAATGTGCTTCTC 130

Db         62 YAMWGTYKKKAMCRTKTKKKKKKGYMWMWYWGWRRSYMAMWTRTWTGYAYYRSMMYWWRY 121

Qy        131 CGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACTGATGTGGTTT 190

Db        122 RCWKKKAYYRKTTCYSSKGWTWWKRWKKAWTTWWWKKTYYWAATRYWWMMCWTKRWRASW 181

Qy        191 TTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATAACAAT---AC 247

Db        182 WYCWWWGKARKWSTWRKSRSYASARSAKRCCYSCSWGAMSWKYMWRMWRWRGWATGAGMK 241

Qy        248 ATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTATTTTATA 307

Db        242 AWRASCMMRRKYAGKSKTSYKSMWMCWTRSWKYCYTKARWTGYYCYRKGGMWGKRGRWYA 301

Qy        308 ATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCGAATTTTT 367

Db        302 SKKYMWKRWWWCWARMYRYSTGTRASMWWRRWYYTMMMKWWKYAWARAAWRWWAMWWAWR 361

Qy        368 CAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAAT 413

Db        362 RACAAAATATAATTATTTATGGTACAATTCTTGTACTTTAGCAAAT 407


Search completed: May 12, 2006, 03:51:30
Job time : 149.274 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:31:18 ; Search time 647.329 Seconds
(without alignments)
12915.142 Million cell updates/sec

Title: US-09-423-093B-3_COPY_6313_7323
Perfect score: 1011
Sequence: 1 atgaataataaaaaggtttt.....catgtgatggaaataaatga 1011

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1011	100.0	12441	3	US-09-423-093A-3	Sequence 3, Appli
	2	61.6	6.1	1014	9	US-10-496-207-31	Sequence 31, Appl
	3	56.6	5.6	22080	3	US-09-423-093A-4	Sequence 4, Appli
	4	53.6	5.3	2731748	7	US-10-297-465A-1	Sequence 1, Appli
	5	52	5.1	16258	7	US-10-257-166-120	Sequence 120, App
	6	51.6	5.1	3673778	6	US-10-312-841-1	Sequence 1, Appli
	7	51.2	5.1	6013	7	US-10-221-613-226	Sequence 226, App
c	8	50.6	5.0	3673778	6	US-10-312-841-1	Sequence 1, Appli
	9	50.4	5.0	1116	6	US-10-369-493-45333	Sequence 45333, A
c	10	50	4.9	42027	8	US-10-417-375-58	Sequence 58, Appl
	11	49.6	4.9	1173	6	US-10-369-493-24780	Sequence 24780, A
	12	49.6	4.9	15548	6	US-10-311-455-2127	Sequence 2127, Ap
	13	49.4	4.9	960	5	US-10-198-846-6381	Sequence 6381, Ap
c	14	49.4	4.9	50000	8	US-10-706-635-25	Sequence 25, Appl
	15	49.2	4.9	3673778	6	US-10-312-841-2	Sequence 2, Appli
	16	49	4.8	8056	8	US-10-473-126-386	Sequence 386, App
	17	49	4.8	640681	3	US-09-790-988-1	Sequence 1, Appli
	18	48.6	4.8	1194	7	US-10-282-122A-16489	Sequence 16489, A
	19	48.2	4.8	15674	6	US-10-311-455-336	Sequence 336, App
	20	48.2	4.8	15674	6	US-10-240-485-30	Sequence 30, Appl
	21	48	4.7	14950	6	US-10-311-455-1230	Sequence 1230, Ap
c	22	47.8	4.7	3673778	6	US-10-312-841-2	Sequence 2, Appli
	23	47.6	4.7	8056	8	US-10-473-126-240	Sequence 240, App
c	24	47.4	4.7	119036	8	US-10-741-600-17721	Sequence 17721, A
	25	47.2	4.7	7498	6	US-10-311-455-230	Sequence 230, App
c	26	47	4.6	631	3	US-09-814-353-5993	Sequence 5993, Ap
c	27	47	4.6	631	3	US-09-814-353-12272	Sequence 12272, A
	28	47	4.6	1410	8	US-10-706-635-33	Sequence 33, Appl
	29	47	4.6	3032	8	US-10-357-930-24019	Sequence 24019, A
	30	47	4.6	3032	8	US-10-357-930-29932	Sequence 29932, A
c	31	47	4.6	3115	3	US-09-764-855-320	Sequence 320, App
c	32	47	4.6	3115	5	US-10-072-349-320	Sequence 320, App
c	33	47	4.6	3454	7	US-10-115-635-219	Sequence 219, App
c	34	47	4.6	4184	3	US-09-374-046A-57	Sequence 57, Appl
c	35	47	4.6	4184	7	US-10-616-263-57	Sequence 57, Appl
	36	47	4.6	50000	8	US-10-706-635-24	Sequence 24, Appl
	37	46.8	4.6	1164	6	US-10-349-680-157	Sequence 157, App
	38	46.6	4.6	5020	8	US-10-473-126-100	Sequence 100, App
c	39	46.6	4.6	32392	8	US-10-706-635-27	Sequence 27, Appl
	40	46.6	4.6	50000	8	US-10-706-635-23	Sequence 23, Appl
	41	46.4	4.6	11996	6	US-10-240-485-45	Sequence 45, Appl
	42	46.2	4.6	16545	6	US-10-311-455-23	Sequence 23, Appl
c	43	46.2	4.6	50000	8	US-10-706-635-23	Sequence 23, Appl
c	44	46	4.5	325	4	US-09-925-065A-201658	Sequence 201658,
c	45	46	4.5	640681	3	US-09-790-988-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-423-093A-3
; Sequence 3, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12441
; TYPE: DNA
; ORGANISM: Salmonella enterica
US-09-423-093A-3

Query Match 100.0%; Score 1011; DB 3; Length 12441;
Best Local Similarity 100.0%; Pred. No. 2.2e-212;
Matches 1011; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6313 ATGAATAATAAAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGT 6372

Qy         61 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6373 TTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAA 6432

Qy        121 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6433 TGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACT 6492

Qy        181 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6493 GATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATA 6552

Qy        241 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6553 ACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTA 6612
```

```
Qy        301 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6613 TTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCG 6672

Qy        361 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6673 AATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACG 6732

Qy        421 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6733 GTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTAT 6792

Qy        481 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6793 AAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCT 6852

Qy        541 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6853 GCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGT 6912

Qy        601 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6913 AATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTGGG 6972

Qy        661 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6973 TTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCT 7032

Qy        721 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7033 TCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTA 7092

Qy        781 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7093 TTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGAC 7152

Qy        841 CCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7153 CCTCTTTCGGAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTT 7212

Qy        901 CGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7213 CGTAAGCATTTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTG 7272

Qy        961 GTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGATGGAAATAAATGA 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7273 GTTAGTGAGATTGAAATGGTACTGACAGAGGCATGTGATGGAAATAAATGA 7323
```

RESULT 2
US-10-496-207-31
; Sequence 31, Application US/10496207
; Publication No. US20050079168A1
; GENERAL INFORMATION:
; APPLICANT: Karlyshev, Andrey
; APPLICANT: Wren, Brendan
; TITLE OF INVENTION: Virulence Genes, Proteins, and Their Use
; FILE REFERENCE: GJE-6947
; CURRENT APPLICATION NUMBER: US/10/496,207
; CURRENT FILING DATE: 2004-05-19
; PRIOR APPLICATION NUMBER: PCT/GB02/05212
; PRIOR FILING DATE: 2002-11-18
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Yersinia pestis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1014)
US-10-496-207-31

Query Match 6.1%; Score 61.6; DB 9; Length 1014;
Best Local Similarity 49.7%; Pred. No. 0.0028;
Matches 157; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

```
Qy        521 ATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCG 580
              || |||| | ||  ||  | |      || | | || | |    ||       | | |
Db        542 ATAAGAATTTTGACATTGCAGTGAAAGCAATATCAAAAACACCTCACCTCGAATTATCAG 601

Qy        581 TTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTG 640
              || |   |||     | |  | |  |  ||||    | || |  | ||   ||||  ||
Db        602 TTGTAGGTGGAGGGGCATTCACTAGTAAAGAACTGTCACTACTGAATCACTATTTACCTG 661

Qy        641 AACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCT 700
              ||      |   || ||  | |  |||  | | |||       |  ||||||
Db        662 GTCGCTACCATGGATTAGGTCGCCTTAGTGATGAGGCTTTGAATGAGGCATATAATTCAG 721

Qy        701 CGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCA 760
              | |  |   | ||| ||| |||   ||||| |||||||| | ||  || | || |  |
Db        722 CTTATGCGCTGCTTTACCCATCTAGCTATGAGGGTTTTGGTATTCCAATATTAGAAGCGA 781

Qy        761 TGGCCTGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAG 820
              ||       ||   ||| ||| || ||  |       || ||  | || ||||| || || |
Db        782 TGAGTGCAGGATGTCCCGTAATATCTGTTAATGTATCTTCTATACCTGAGGTCGCAGGTG 841

Qy        821 ATGCAGCGATTCTTGT 836
              |||| || ||  | ||
Db        842 ATGCCGCTATATTAGT 857
```

RESULT 3
US-09-423-093A-4
; Sequence 4, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 22080
; TYPE: DNA
; ORGANISM: Salmonella enterica
US-09-423-093A-4

Query Match 5.6%; Score 56.6; DB 3; Length 22080;
Best Local Similarity 46.8%; Pred. No. 0.14;
Matches 296; Conservative 0; Mismatches 319; Indels 18; Gaps 3;

```
Qy        225 GAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTC 284
              || |||    |  |||||  |||||||| |||  | | |  |||     |||| ||
Db      15690 GAGAAACCAAACGATAACCGTACATGATATAAGACCTTTTTATTATCCAGATAGTT---- 15745

Qy        285 TCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAA 344
              || || |||   | | || |   ||| || ||||         ||    |    | |
Db      15746 --TTATTCAGAAAGTGTATTTTCGCTTTTTATTAAAAATGTCCGTTAAGCGATGTAAGCA 15803

Qy        345 AATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCC 404
              | || || || ||  ||      |||| || || |    |   |   |||| |
Db      15804 TGTTTTAACGGTATCTTATACCGTTAAAGATAGCATTGCTAAAACTTATAATGTAGATAG 15863

Qy        405 TAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACC 464
              | | ||||||       | |||||         |||| |   |||  | ||| | | ||
Db      15864 TGAGAAAATATCAGTAATTTATAA---------TAGTGTTAATAAATCTGATTTTATACA 15914
```

```
Qy        465 ATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGA 524
Db      15915 AAAAAAAGAAAAAGAGAATTACTTTTTAGCTGTTGGTGCAAGTTGGCCACATAAAAATAT 15974

Qy        525 GAAGTGTGTTAT---ATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGT 581
Db      15975 TCATTCATTCATAAAAAATAAAAAAGTTTGGTCTGACTCTTATAATTTAATTATTGTATG 16034

Qy        582 TTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGA 641
Db      16035 TGGTCGTACTGACTATGCAATGTCTCTCCAACAAATGGTCGTTGATCTGGAACTAAAAGA 16094

Qy        642 ACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTC 701
Db      16095 TAAAGTGACTTTTTTACATGAAGTCTCATTTAATGAATTAAAGATTTTATATTCTAAAGC 16154

Qy        702 GTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCAT 761
Db      16155 CTACGCGCTTGTTTATCCATCTATTGATGAGGGTTTTGGTATACCTCCTATTGAAGCGAT 16214

Qy        762 GGCCTGTGGTATTCCTGTATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGA 821
Db      16215 GGCATCAAATACTCCAGTTATAGTGTCCGATATACCAGTATTTCATGAAGTGTTAACCAA 16274

Qy        822 TGCAGCGATTCTTGTCGACCCTCTTTCGGAAGA 854
Db      16275 TGGTGCATTATATGTGAATCCGGATGATGAAAA 16307
```

RESULT 4
US-10-297-465A-1
; Sequence 1, Application US/10297465A
; Publication No. US20040142413A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Setubal, Joao
; APPLICANT: Medianis, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465A
; CURRENT FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465A-1

Query Match 5.3%; Score 53.6; DB 7; Length 2731748;
Best Local Similarity 49.0%; Pred. No. 5.6;
Matches 143; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

```
Qy        670 GAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTTATAT 729
Db     847312 GATAGCGATTTGGTCGCCCTCTATGGGCTGTGTGCGTTGTTCGTGTTCCCATCGCGTTTG 847371

Qy        730 GAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTGTATTAACTTCT 789
Db     847372 GAAGGTTTCGGGCTTCCCGTGTTGGAGGCGATGGCCTGTGGCGCACCCACATTGTGCAGT 847431

Qy        790 CTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCGACCCTCTTTCG 849
Db     847432 GATGCGGCTAGCTTGCCAGAGGTCGCGGGGCGTCAGGACATCCTGTTTCCACCTGAAGAT 847491

Qy        850 GAAGATGCTATTACTAAAGGAATTTCGAGGTTAATTAATGATTCTGAACTTCGTAAGCAT 909
```

```
Db     847492 CATGTTGCCTTGAGTGCTCGCATGTCCCAGGTGCTTGATGATCCACAGTGGTTACAGTCG 847551

Qy        910 TTAATCCAAAAGGGGCTTTTGCGGGCAAAGAGGTTCAATTGGCAAAACGTGG 961
Db     847552 CTGCGTGATTATGGTTTGCGCCGTGCTACAGCATTCACTTGGGATCGCGTGG 847603
```

RESULT 5
US-10-257-166-120
; Sequence 120, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 120
; LENGTH: 16258
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-120

Query Match 5.1%; Score 52; DB 7; Length 16258;
Best Local Similarity 47.3%; Pred. No. 1.3;
Matches 157; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

```
Qy        182 ATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATAA 241
Db       2140 ATATGTTTTTTATATGTATATATATTATATATTTTATATTTTTATATGTTTTTAGTATAT 2199

Qy        242 CAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTAT 301
Db       2200 ATATATATTTTTTATATATATTTTATATATATATATTTATATTTTATTTTTATATATATG 2259

Qy        302 TTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCGA 361
Db       2260 TATATATATATATATTATATATATTATATATATATATTATATATATATTTTATATATATA 2319

Qy        362 ATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACGG 421
Db       2320 TTATATATATATATATTTATTTTTTATATATATAGGTAGAAATATATATATTATATATAT 2379

Qy        422 TATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTATA 481
Db       2380 TATATATATATATATATAGTTTATATATTTTTTATATATATATATTTTATATATATATTA 2439

Qy        482 AATATTTGCTATGTGTAGGAAACAGAAAAACT 513
Db       2440 TATATATATAGTTTATTTTATATATATATAGT 2471
```

RESULT 6
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841

```
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3294164)
US-10-312-841-1

  Query Match             5.1%;  Score 51.6;  DB 6;  Length 3673778;
  Best Local Similarity   44.2%;  Pred. No. 18;
  Matches  257;  Conservative    0;  Mismatches  324;  Indels    1;  Gaps    1;

Qy         60 TTTTACTGATGAAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAA 119
              |||||  ||  |  |  |  |    ||| | |||||| ||   |     |||||| |
Db    1714987 TTTTATATATATATATATTTTATATATATTTTATATATATATATTTTATATATATATT 1715046

Qy        120 ATGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAAC 179
              | |   | |              ||      ||      | |||| | | |   | | | |
Db    1715047 TTATATATATATATATTTTATATATATATTTTATATATATATATTTTATATATATATA 1715106

Qy        180 TGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAAT 239
              |  | |   | |  ||  |   ||||||  |    | | ||  |    | |  ||    ||
Db    1715107 TTTTATATATATATTTTATATATATATTTTATATATATATTTTATATATATATTTTATAT 1715166

Qy        240 AACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTT 299
              |   ||     || || ||  | |  || || ||  ||  |       |||||      |
Db    1715167 ATATATTTTATATATATATATTTTATATATATATTTTATATATATATTTTATATATAT 1715226

Qy        300 ATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTC 359
              |||||||| | |||  || |      || || | |   || | | |||| ||   | |
Db    1715227 ATTTTATATATATATTTTATATATATATTTTATATATATATTTTATATATATATTTTATA 1715286

Qy        360 GAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCAC 419
              |  | ||  | | | | | |      ||  |    | ||| |      | | |||   |
Db    1715287 TATATATTTTATATATATATTTTATATATA-TATTTTATATATATATTTTATATATATAT 1715345

Qy        420 GGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTA 479
              | ||||    | | |  | ||| || |    |||| |     |  || ||   ||
Db    1715346 ATTTTATATGTATATTTTATATATATATATTTTATATGTATATTTTATATATATATTTTA 1715405

Qy        480 TAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATC 539
              ||  ||| |  ||| |||| |  | | | |   | | |   | |   |    |||||
Db    1715406 TATGTATATTTTATATGTATATTTTATATATATATTTTATATGTATATTTTATATATATA 1715465

Qy        540 TGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTG 599
              |    |    | |     |||  ||  ||  ||| | |  | | ||    |  | | |
Db    1715466 TTTTATATGTATATATTTTATATATATATATATATTTTATATATATATTTTATTAG 1715525

Qy        600 TAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGA 641
              |  || |||   | ||   | ||      | |||  || |
Db    1715526 TTTTGTTTTTTTAGAAAATTTGATTAGTATAGATTTTGGTAA 1715567


RESULT 7
US-10-221-613-226
; Sequence 226, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, ALexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
;  FILE REFERENCE: 5013.1004
;  CURRENT APPLICATION NUMBER: US/10/221,613
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: PCT/EP01/02945
;    DE 10013847.00
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-03-15
;    2000-03-15
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 226
;   LENGTH: 6013
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-226

  Query Match             5.1%;  Score 51.2;  DB 7;  Length 6013;
  Best Local Similarity   42.9%;  Pred. No. 1.2;
  Matches  254;  Conservative    0;  Mismatches  338;  Indels    0;  Gaps    0;

Qy        158 ATATTTTTCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTT 217
              || |||||  |  ||  ||    || | | |||| |   |   ||||      | |||
Db         68 ATTTTTTTTAGGAAATTAATTAATATATAGATTTTATGAATTTTTATAAGTAGTTAATTT 127

Qy        218 TTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATA 277
              ||| |   ||||| ||  ||  ||       ||||   || |  | |  | | || | |||
Db        128 TTTATAAAAAAAAATTAGTATGAAATGGGTATGAAAAAATTTTTGTAAAATAAATTTATA 187

Qy        278 ATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAG 337
              || | | |   | || |   |    ||  |||||| || |  |   |    | || ||    |
Db        188 ATACGTATAAATAAAATAATGTGAAATAGTATTTTAAAAAATTGAATGTTTTTAATTTAT 247

Qy        338 CATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTG 397
                | | | | |     | | | | || |  | |   || ||  ||| || ||    |
Db        248 TGATTTAGGTTTAATATATATAGGAATTATATATATTTAGTATGTTAGTATTGTTTATTA 307

Qy        398 TAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATG 457
              |       | |    |   | | |   || | |||       |  |  || |    |
Db        308 AATTTATAATTTGTTGATATATGTTTATTATTAGGGGTATATTTTAAAATATTTTAAAAT 367

Qy        458 TAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATA 517
              |  ||        ||  ||    ||  | |||   ||     |     || || |  ||
Db        368 TTGAAAAGAATAAAATATAAGATTTAAATTATATAAAATTTATTATAGTAAGAAATTTTA 427

Qy        518 AGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAAC 577
              |    |||||    ||| | | |   |     | |   | || |  |   |    ||  |
Db        428 ATTGAGAGAAAAAAGTTTTTTTTAGTTGGTTGAGATATGTTAATTTTAAAATTTTATTAT 487

Qy        578 TCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTAA 637
              |     ||| |  | ||     |  |   ||     |  | ||     || ||  ||| |
Db        488 TGAAGATTATTATTTATTTAAATTTTTGATTTAGTGATTTTATTTAGAAATATTTTTTTA 547

Qy        638 GTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGG 697
              |      |  || ||| |||    || || | | | |   |    | ||   ||
Db        548 TTTTTTTTTAAACGTTGTTTTATTTTGTTTTTTAGATTGGAGTGTAGTGGTGCGATTTTT 607

Qy        698 GCTCGTTAGGATTAGTTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGT 749
              | |  ||  ||||    ||      ||||   ||   |  |  |||  | |
Db        608 GTTTATTGCGATTTTCGTTTTTAGGTTTAAGCGATTTTCGTTTTTTAGATAT 659


RESULT 8
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
```

```
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3294164)
US-10-312-841-1

  Query Match             5.0%;  Score 50.6;  DB 6;  Length 3673778;
  Best Local Similarity   45.5%;  Pred. No. 29;
  Matches  179;  Conservative    0;  Mismatches  214;  Indels    0;  Gaps    0;

Qy        238 ATAACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGG 297
              |||   ||| |  || || || |  |  || || |  |||  |         | ||
Db    1715170 ATATATATAAAATATATATATAAAATATATATATAAAATATATATATAAAATATATATAT 1715111

Qy        298 TTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTT 357
              | | ||||  | || |||| |      |   | | |  ||||||||||| || |  |
Db    1715110 AAAATATATATATATAAAATATATATATATAAAATATATATAAAATATATATATAT 1715051

Qy        358 TCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTC 417
              || | |  | |  ||| |    ||         | | || |   ||   | |||
Db    1715050 ATAAAATATATATATAAAATATATATATATATAAAATATATATAAAATATATATATAT 1714991

Qy        418 ACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGC 477
              |  ||||||      ||     | ||| |||      || || |   | |  || ||  |
Db    1714990 AAAATATATATATATATAAAATATATATATAAAATATATATATACAAAATATATATATAC 1714931

Qy        478 TATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATA 537
              | | |||| |   |  | ||  || | | | |  | || |   | | |    ||||
Db    1714930 AAAATATATATATAAAATATATATAAAATATATATATAAAATATATATATATAAAATATA 1714871

Qy        538 TCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCT 597
              | |    |    | |    | ||  ||  |  |  |  | |     | | | |  ||    |
Db    1714870 TATAAAATATATACATATAAAATATATATACTATATATATAAAATATAAATAATATATAT 1714811

Qy        598 TGTAATGATTTAGAAAAACTAATAATACAACAT 630
              ||  || || | ||| ||   ||| || ||
Db    1714810 ATAAAATATATATATAAAATATATATAAAATAT 1714778


RESULT 9
US-10-369-493-45333
; Sequence 45333, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 45333
;   LENGTH: 1116
;   TYPE: DNA
;   ORGANISM: Pyrococcus abyssi
US-10-369-493-45333

  Query Match             5.0%;  Score 50.4;  DB 6;  Length 1116;
  Best Local Similarity   50.4%;  Pred. No. 0.85;
  Matches  123;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy        599 GTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGTTCTTTG 658
              | || | | |  | |  | |  || | | |  ||  |    ||    ||  | ||| |||
Db        692 GAAAAGTTCTCAAGATGCAAGCAAAAAAGCTGGGAGTGGAAGATAAAGTCGAATTCCTTG 751

Qy        659 GGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAGTTTTCC 718
              | ||  |      |  ||||  ||   |  ||||||||         || ||| ||  | |
Db        752 GATTTATAAGCCAGGAAGAGCTAATCTTACTATATAAGAAAAGCGAAGTATTCGTCCTAC 811

Qy        719 CTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTATTCCTG 778
              | ||  |    || | ||| ||  ||  | || | || | ||||||  | |   | || |
Db        812 CATCACTCAGCGAGGCTTTCGGGATAGTTCTACTCGAAGCCATGGCGAGCGAGGTCCCAG 871

Qy        779 TATTAACTTCTCTAACTTCATCATTGCCAGAGGTGGCTGGAGATGCAGCGATTCTTGTCG 838
              |  ||    |  |        | | |||||| |    ||||||||||  ||  | |||
Db        872 TGATAGGAACCTCAGTGGGTGGAATTCCAGAGATAATAGGAGATGCAGGAATAATAGTCC 931

Qy        839 ACCC 842
              ||
Db        932 CTCC 935


RESULT 10
US-10-417-375-58/c
; Sequence 58, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001600
;  CURRENT APPLICATION NUMBER: US/10/417,375
;  CURRENT FILING DATE:  2003-04-15
;  NUMBER OF SEQ ID NOS: 176
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 58
;   LENGTH: 42027
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-417-375-58

  Query Match             4.9%;  Score 50;  DB 8;  Length 42027;
  Best Local Similarity   45.4%;  Pred. No. 5.3;
  Matches  179;  Conservative    0;  Mismatches  215;  Indels    0;  Gaps    0;

Qy        153 AGTCAATATTTTTCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACC 212
              | |  ||| |||         | | |    || |  ||| | || |    ||||   |
Db      30167 AATTTATAATTTATAATATATTATATAAAATATAATTTATATTACATAAAATATATAATT 30108

Qy        213 ACTTTTTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAA 272
              |  | | ||  | |||  |   | | | ||||  |  || ||||  || |   ||  |
Db      30107 TATAATATATTATATAAACATATAATTTATAATATAATATATAAAATATATAACTTATAT 30048

Qy        273 TGATAATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAG 332
              |    |           ||||  |  |  | |   || | ||||||  |       |
Db      30047 TATATAAAATATATAATTTATAATATATTATATAAAATATATAATTTATAATATATTATA 29988

Qy        333 AAAAGCATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTC 392
              | |  |||||| |||||     ||   ||   |  | || | |  |||          |
Db      29987 TATAATATATAATATATTATATATTATATATAATATATAATATATTATATATTATATATA 29928

Qy        393 AGGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGC 452
              |  | |||    ||||| | ||   |   ||||||||    ||| || | ||| ||||
Db      29927 ATATATAATTTATAATATATTATATATAATATATAATATATTATATATTATATATAATAT 29868

Qy        453 CGATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAAC 512
              ||   | | |||  ||  ||       | ||||    |||       | |   | | |
Db      29867 ATATTATATAACATATAAAATATATATAATATATAATATATAAAATATAGATTTATATAA 29808

Qy        513 TCATAAGAATGAGAAGTGTGTTATATCTGCCTTT 546
              ||  |    || |  ||||| |  ||||
Db      29807 AGGGGAGTTTATTTGGTATTATATATATATCTTT 29774
```

```
RESULT 11
US-10-369-493-24780
; Sequence 24780, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 24780
;   LENGTH: 1173
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
US-10-369-493-24780

  Query Match             4.9%;  Score 49.6;  DB 6;  Length 1173;
  Best Local Similarity   50.0%;  Pred. No. 1.3;
  Matches  155;  Conservative    0;  Mismatches  149;  Indels    6;  Gaps    1;

Qy        479 ATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATAT 538
              | |||     |  ||| ||| ||||     || |  ||| |||        ||| | | |||
Db        623 AAAAAATGATTTTATTTGTTGGAAGATTAACATATCAAAAGGGAATAGAGTATTTAATAA 682

Qy        539 CTGCCTTTGCCAAA------GCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTA 592
              ||  |  | |||       | |    | || ||  | ||    | |    | ||||
Db        683 GAGCAATGCCAAAAATTCTTGAAAGACATAATGCAAAATTAGTTATTGCAGGTTCTGGAG 742

Qy        593 ATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTTAAGTGAACGTGTAAAGT 652
              ||     |  || |||| ||| |  |        || | |||| | ||  |    |||  ||
Db        743 ATATGAGGGATTATTTGGAAGATTTGTGCTATCAGCTTGGTGTTAGGCATAAGGTAGTGT 802

Qy        653 TCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATATAAGGGCTCGTTAGGATTAG 712
              | || || || ||   || | |    |||  |    | |||||      ||   |  | |
Db        803 TTTTAGGATTTGTTAATGGAGATACGTTAAAAAAACTCTATAAATCTGCGGATGTTGTTG 862

Qy        713 TTTTCCCTTCTTTATATGAAGGTTTTGGATTACCTGTAGTGGAGGGCATGGCCTGTGGTA 772
              |  | || ||  | |||||     |||||  ||  ||   | ||||  |||||   ||| |
Db        863 TAATTCCATCAGTTTATGAGCCATTTGGTATAGTTGCTTTAGAGGCAATGGCTGCTGGCA 922

Qy        773 TTCCTGTATT 782
              || ||| |
Db        923 CACCAGTAGT 932


RESULT 12
US-10-311-455-2127
; Sequence 2127, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2127
;   LENGTH: 15548
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2127

  Query Match             4.9%;  Score 49.6;  DB 6;  Length 15548;
  Best Local Similarity   45.8%;  Pred. No. 4.2;
  Matches  284;  Conservative    0;  Mismatches  329;  Indels    7;  Gaps    3;

Qy         11 AAAAGGTTTTGATGGATATTAGTTGGTCTAATAAAGGGGGGATTGGACGTTTTACTGATG 70
              |||     | |   | ||||   |  ||  ||||||         ||||   | ||    ||
Db       1421 AAATATATATATGGAATATATATATGTGGAATAAATATATATATGGAATATATATATGTG 1480

Qy         71 AAATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAATGTGCTTCTC 130
              ||||  || |    |     |||||| |  |||   | |||||| |     |   | ||
Db       1481 GAATTAATATATATATGGAATATATATATGTGGAATTAATATATATATGTAATATATATA 1540

Qy        131 CGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACTGATGTGGTTT 190
              | |     ||     |  | | |  |||| | |      || | |  | ||||  | |
Db       1541 TGTGGAATTAATATATATATGTAATATATATATGTGGAATTAATATATAT-ATGTAATAT 1599

Qy        191 TTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATAACAATACATG 250
              | |   |||   || |  |    | |  | |     |   |    |||||   |||  |
Db       1600 ATATATGTGGAATTAATATATATATGTAATATATATATGTGGAATTAATATATATATGTA 1659

Qy        251 ATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTATTTTATAATT 310
              || || || || |||| ||||   |||  |    | | | || | ||   |  ||||  |
Db       1660 ATATATAT-ATGTGGAATTAATATATATATGTAATATATATATGTGGAATTAATATATAT 1718

Qy        311 TTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCGAATTTTTCAA 370
              | ||||| |     || | |  ||   |||||    || || |  |  | | ||
Db       1719 ATGTAATATATATATGTGGAATTAATATATATATGTAATATATATATGTGGAATT----- 1773

Qy        371 AAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACGGTATATAATG 430
              || | | |   |||  ||    | ||| ||    ||   | ||||      |||||   |
Db       1774 AATATATATATGTAATATATATATGTGGAATTAATATATATATAGAATATATATATGTGG 1833

Qy        431 GGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTATAAATATTTGC 490
              ||     | |     || |  ||||  ||    |  || | |  |||| ||| ||
Db       1834 AATTAATATATATGGAATATACGTATGTGGAATATATATATGTGGAATATACGTATGTGG 1893

Qy        491 TATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCA 550
              | | ||    | | ||  ||  | || | |  | |||  || | | | ||
Db       1894 ATTATATATATGTGGAATATACGTATGTGGATTATATATATGTGGAATATACGTATGTGG 1953

Qy        551 AAGCAGATATTGATCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAG 610
              |         ||   |  |  ||| || |||   |     || ||   || | |
Db       1954 ATTATATATATGTGAAATATACGTATGTCGATTATACGTATGGGTTTTATATATATATGG 2013

Qy        611 AAAAACTAATAATACAACAT 630
              | |  ||   ||  || ||
Db       2014 GATATATACGTATGGAATAT 2033


RESULT 13
US-10-198-846-6381
; Sequence 6381, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
```

```
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6381
;   LENGTH: 960
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1, 8, 66, 71, 79, 81, 84, 85, 86, 88, 117, 119, 125, 134,
;   LOCATION: 149, 156, 162, 166, 172, 179, 180, 194, 230, 249, 250, 253,
;   LOCATION: 268, 271, 273, 274, 275, 276, 277, 286, 291, 299, 312, 333,
;   LOCATION: 359, 366, 374, 393, 404, 406, 409, 411, 422, 424, 427
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 428, 433, 450, 465, 466, 468, 471, 487, 510, 512, 513, 516,
;   LOCATION: 517, 518, 528, 530, 534, 539, 553, 555, 559, 575, 580, 588,
;   LOCATION: 590, 610, 625, 626, 642, 645, 656, 658, 659, 675, 685, 686,
;   LOCATION: 690, 700, 702, 706, 716, 731, 745, 748, 756, 798, 802
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 807, 808, 819, 820, 821, 830, 838, 840, 844, 845, 849, 861,
;   LOCATION: 871, 874, 875, 891, 892, 895, 898, 900, 901, 902, 904, 907,
;   LOCATION: 917, 923, 925, 926, 927, 929, 955
;   OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6381

  Query Match             4.9%;  Score 49.4;  DB 5;  Length 960;
  Best Local Similarity   38.7%;  Pred. No. 1.3;
  Matches  185;  Conservative    0;  Mismatches  293;  Indels    0;  Gaps    0;

Qy        216 TTTTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGA 275
              ||    ||  ||    |     ||||  ||| |     |   | |     |||| | |
Db        407 TTNANTTTTAAATTTNTNTTNNATAANTATAAAATTATTTTTTNAAAATTATTTTATTNN 466

Qy        276 TAATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAA 335
              |   |   |  ||| ||  |     |    |||| ||| | |                | ||
Db        467 TNTTNTTTTAATTTATAGAANTAAAAAAAATAAATTTTTTAATNTNNAANNNAAAAAAAT 526

Qy        336 AGCATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGG 395
              | | || |     |    |   ||| |||| |     |  |  |    |
Db        527 TNTNTTTNAATTNAAGGTTTTAAAAANANTTTNAAAAAATTTTTTTTTNTTTTNAGAAAA 586

Qy        396 TGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGA 455
                ||    |||    ||   |  || | ||        | |  | |     ||    |
Db        587 ANANAAATAAAATTTTTTATAAANTATAAAAAAAAAAAANNTTTTTTTTTAAATATNTTNA 646

Qy        456 TGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCA 515
              |  ||||   |    |||    || |||  |||  |||     |  ||| | || |  |
Db        647 TAAAAAAAANTNNTTTTTTATTTAAAAANTTTTTTTATNNAAANAAAAAAAAAANANTTGN 706

Qy        516 TAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATAAA 575
              | | |||   || | | || | |          ||   |  |  | |    | || |||
Db        707 TTAAAATTTNAAATTTTTTTTTTTTNATTAAAAAAAATTNAANTTATTAANTTTTAAAAAA 766

Qy        576 ACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATGGTTT 635
              |  ||   |     | |     | | |  |  ||||    ||||          |||
Db        767 TTTTTTTAAAAAAAAAATTTTGGTTTAATAANTTANAAAANNTTTAATTTTTNNNTTTTT 826

Qy        636 AAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTATAT 693
                ||  |  |   |  ||    || || | | |||||  ||     |    ||||
Db        827 TTANAAAAATTNANTATNNTTANAATTAGTTTTTNAAAAAATTTNTANNTTTAAATAT 884


RESULT 14
US-10-706-635-25/c
; Sequence 25, Application US/10706635
; Publication No. US20050014263A1
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XCZ1
;  CURRENT APPLICATION NUMBER: US/10/706,635
;  CURRENT FILING DATE:  2003-11-12
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 09/662,254
;  PRIOR FILING DATE: 2000-09-14
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-10-706-635-25

  Query Match             4.9%;  Score 49.4;  DB 8;  Length 50000;
  Best Local Similarity   47.1%;  Pred. No. 7.8;
  Matches  152;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy        117 AAAATGTGCTTCTCCGCTGGCCCCATTAGGTTTAGCAGTCAATATTTTTCTGCGAAAGAA 176
              || || ||||  | |          |||| |||||    | |||| | |        ||  |
Db      42125 AATATTTGCTCATACAGACATAGAATTATGTTTATATTTAAATAATATAGAAATATAAAGA 42066

Qy        177 AACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCAT 236
              |  |          |  ||  | |||    || |  |  | |       ||  || || || ||
Db      42065 AGATATAAATTTAATTAATTCTGATAGTAATATATTATTGAAACATTATAATAATTTAAT 42006

Qy        237 AATAACAATACATGATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAG 296
              ||||  |    || ||  | || |  |   |   | |   |||   || |  | ||| |
Db      42005 AATAGAATGTGATAATGAATATAAAATAATTAATAGTATAAATATTTCACAATATAAAAA 41946

Qy        297 GTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGT 356
              ||  | |||||   ||||||  |          || |        |||| | || |    |
Db      41945 CATAACTAATAATAGAATAATATATTATAAACATAAATGTTTTAATAATAAATCTTGTTT 41886

Qy        357 TTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGT 416
              |    || ||||| ||||   |   ||  ||  | |  | |||     || ||||||  |
Db      41885 GTTTGTTTATTCAATAGAATATAATATATTATATTTATTTTTAATAAAAAAAAAAATATT 41826

Qy        417 CACGGTATATAATGGGGTATCTA 439
              |  ||  |  |    |||| ||
Db      41825 AAGTGTTAAGGAATATGTATATA 41803


RESULT 15
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (379615)
US-10-312-841-2

  Query Match             4.9%;  Score 49.2;  DB 6;  Length 3673778;
  Best Local Similarity   48.3%;  Pred. No. 59;
  Matches  138;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;
```

```
Qy        215 TTTTTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATG 274
              | |||| || || ||| || ||   || | |  ||  |  |||| ||    | ||||||
Db    1970869 TATTTTTTTAGAGAAATTTTATGTAAATATTTGATTTTTAAAATTATGGATAATTAAATT 1970928

Qy        275 ATAATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAA 334
              | |||        ||||||   |       ||   | | |||| |     |  | ||  |
Db    1970929 TTGATTATAAAAATTTTAATGAATATGAAAATTTATATTTAATGAGAATAGAATTTAATA 1970988

Qy        335 AAGCATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAG 394
              |   ||||    || ||| | |    ||  ||   ||| | | ||||  |    |    |
Db    1970989 AGTGATATTTGTTAATTAAAATAAATAAGATTAATAAATATATAATATAAATGTATGAAT 1971048

Qy        395 GTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCG 454
              |||||||   || |||  |     |     | | ||   | || | | | |||||   |
Db    1971049 GTGTAAATGATATTAATTTGAAAATATAAAAGAAAATGTTAACGAATATATTTTTGGATT 1971108

Qy        455 ATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGG 500
              |   |  | ||   | | |  | | | | |    | | | | ||
Db    1971109 AGTAAGTAGTATATGAGTGAATTTTTTATTTTAGTTTTTTTATTGG 1971154
```

Search completed: May 12, 2006, 05:34:25
Job time : 666.329 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:52:36 ; Search time 515.104 Seconds
(without alignments)
8004.956 Million cell updates/sec

Title: US-09-423-093B-3_COPY_6313_7323
Perfect score: 1011
Sequence: 1 atgaataataaaaaggtttt.....catgtgatggaaataaatga 1011

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/2/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/2/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	51.8	5.1	990	12	US-10-301-480-568794	Sequence 568794,
2	51.8	5.1	990	12	US-10-301-480-1182203	Sequence 1182203,
c 3	48.6	4.8	119036	10	US-10-995-561-13314	Sequence 13314, A
4	46.8	4.6	991	12	US-10-301-480-568792	Sequence 568792,
5	46.8	4.6	991	12	US-10-301-480-568795	Sequence 568795,
6	46.8	4.6	991	12	US-10-301-480-1182201	Sequence 1182201,
7	46.8	4.6	991	12	US-10-301-480-1182204	Sequence 1182204,
c 8	46	4.5	325	7	US-09-925-065A-201658	Sequence 201658,
c 9	46	4.5	336	12	US-10-301-480-290170	Sequence 290170,
c 10	46	4.5	336	12	US-10-301-480-903579	Sequence 903579,
11	46	4.5	6257	17	US-11-011-332A-95	Sequence 95, Appl
c 12	45.6	4.5	420	7	US-09-925-065A-165636	Sequence 165636,
c 13	45.6	4.5	427	12	US-10-301-480-257920	Sequence 257920,
c 14	45.6	4.5	427	12	US-10-301-480-871329	Sequence 871329,
15	45	4.5	990	12	US-10-301-480-568793	Sequence 568793,
16	45	4.5	990	12	US-10-301-480-1182202	Sequence 1182202,
17	45	4.5	998	12	US-10-301-480-582177	Sequence 582177,
18	45	4.5	998	12	US-10-301-480-1195586	Sequence 1195586,
c 19	45	4.5	49979	10	US-10-995-561-13443	Sequence 13443, A
20	44.8	4.4	28536	17	US-11-011-332A-156	Sequence 156, App
21	44.6	4.4	1134	11	US-10-932-182A-166160	Sequence 166160,
22	44.6	4.4	1134	11	US-10-932-182A-166160	Sequence 166160,
c 23	44.4	4.4	416	7	US-09-925-065A-165637	Sequence 165637,
c 24	44.4	4.4	423	12	US-10-301-480-257921	Sequence 257921,
c 25	44.4	4.4	423	12	US-10-301-480-871330	Sequence 871330,
c 26	44.4	4.4	432	12	US-10-301-480-257919	Sequence 257919,
c 27	44.4	4.4	432	12	US-10-301-480-871328	Sequence 871328,
c 28	44.4	4.4	435	7	US-09-925-065A-165635	Sequence 165635,
29	44.4	4.4	1457619	17	US-11-098-686-8739	Sequence 8739, Ap
30	44.2	4.4	1271	11	US-10-301-480-37771	Sequence 37771, A
31	44.2	4.4	1271	12	US-10-301-480-651180	Sequence 651180,
c 32	44	4.4	608	7	US-09-925-065A-661135	Sequence 661135,
c 33	44	4.4	1691140	17	US-11-091-018-1	Sequence 1, Appli
c 34	43.8	4.3	469	7	US-09-925-065A-622649	Sequence 622649,
35	43.4	4.3	867	10	US-10-750-185-55721	Sequence 55721, A
36	43.4	4.3	867	10	US-10-750-623-55721	Sequence 55721, A
c 37	43.2	4.3	545	7	US-09-925-065A-531316	Sequence 531316,
38	43	4.3	677	7	US-09-925-065A-667262	Sequence 667262,
39	43	4.3	677	7	US-09-925-065A-667263	Sequence 667263,
40	43	4.3	677	7	US-09-925-065A-667264	Sequence 667264,
41	43	4.3	677	7	US-09-925-065A-667266	Sequence 667266,
c 42	43	4.3	966	12	US-10-301-480-553375	Sequence 553375,
c 43	43	4.3	966	12	US-10-301-480-1166784	Sequence 1166784,
44	43	4.3	990	12	US-10-301-480-568791	Sequence 568791,
45	43	4.3	990	12	US-10-301-480-1182200	Sequence 1182200,

ALIGNMENTS

RESULT 1
US-10-301-480-568794
; Sequence 568794, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 568794
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-568794

Query Match 5.1%; Score 51.8; DB 12; Length 990;
Best Local Similarity 46.0%; Pred. No. 1.4;
Matches 214; Conservative 0; Mismatches 247; Indels 4; Gaps 1;

```
Qy      158 ATATTTTTCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTT 217
            ||||| | |    | ||| | || | | ||   | |   | | | | ||   |
Db      292 ATATTATATTAAATATAAAATATAATATTATATTATATATTATATTAAATATAACATAAT 351

Qy      218 TTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATA 277
              | |     ||| | ||   | |||| | ||   ||| | | | | ||| |||
Db      352 AGTTTGTATTAAATATATTATATATAATATAATATAATAATGTTATATAATGTAATAATA 411

Qy      278 ATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAG 337
              | | |  | |   | | ||| ||| ||||||||||| |       | || || |
```

```
Db        412 TATTATTTTATAATGTAACATAATTATATTATTTTATAATATATAATTTATTTAAAATAT 471

Qy        338 CATATAAAATAT----TTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCA 393
              ||||| |||||    | | |  |    || | |   | |    |||         | |
Db        472 AATATATAATATAATGTAATATATGATTATATATTGTATTATGTTATATGTAATATATTA 531

Qy        394 GGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCC 453
              | |||    | ||| | || | |         |||    ||| || | ||| ||||
Db        532 TATATAATATATGATATATTAATAATATATATAAATAATTTATATATTATATATAATTTA 591

Qy        454 GATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACT 513
              || ||  |   || ||  |    | || |||| |   | | ||  | | |  | ||
Db        592 TATATATTATATTTTAAAATGTAATTTATATATATAAATTATATATTATAAAATATAAAA 651

Qy        514 CATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATA 573
              |||      | ||           |   ||   ||    | || || ||  ||  |||
Db        652 TATATAGTATATAATATATAATACATTATTTTGTTCATTTTAAATTTTTATATATTTATA 711

Qy        574 AAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTA 618
              |      | ||||       |||| | ||   || ||| | |  ||
Db        712 TATTATATGTTTATATTGTATCCATATATAAATGTAGCATATATA 756


RESULT 2
US-10-301-480-1182203
; Sequence 1182203, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1182203
;   LENGTH: 990
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1182203

  Query Match             5.1%;  Score 51.8;  DB 12;  Length 990;
  Best Local Similarity   46.0%;  Pred. No. 1.4;
  Matches  214;  Conservative    0;  Mismatches  247;  Indels    4;  Gaps    1;

Qy        158 ATATTTTTCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTT 217
              |||||  |   |   |  |||  | || | | ||  |  |   |  | |  | ||   |
Db        292 ATATTATATTAAATATAAAATATAATATTATATTATATATTATATTAAATATAACATAAT 351

Qy        218 TTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATA 277
              | |    |||  |  ||   | ||||  |  ||   |||  | |  | |  ||| |||
Db        352 AGTTTGTATTAAATATATTATATATAATATAATATAATAATGTTATATAATGTAATAATA 411

Qy        278 ATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAG 337
              |  |  |  |  |   |  | ||| ||| ||||||||||| |        | || || |
Db        412 TATTATTTTATAATGTAACATAATTATATTATTTTATAATATATAATTTATTTAAAATAT 471

Qy        338 CATATAAAATAT----TTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCA 393
              ||||| |||||    | | |  |    || | |   | |    |||         | |
Db        472 AATATATAATATAATGTAATATATGATTATATATTGTATTATGTTATATGTAATATATTA 531

Qy        394 GGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCC 453
              | |||    | ||| | || | |         |||    ||| || | ||| ||||
Db        532 TATATAATATATGATATATTAATAATATATATAAATAATTTATATATTATATATAATTTA 591

Qy        454 GATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACT 513
              || ||  |   || ||  |    | || |||| |   | | ||  | | |  | ||
Db        592 TATATATTATATTTTAAAATGTAATTTATATATATAAATTATATATTATAAAATATAAAA 651

Qy        514 CATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAATA 573
              |||      | ||           |   ||   ||    | || || ||  ||  |||
Db        652 TATATAGTATATAATATATAATACATTATTTTGTTCATTTTAAATTTTTATATATTTATA 711

Qy        574 AAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTA 618
              |      | ||||       |||| | ||   || ||| | |  ||
Db        712 TATTATATGTTTATATTGTATCCATATATAAATGTAGCATATATA 756


RESULT 3
US-10-995-561-13314/c
; Sequence 13314, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13314
;   LENGTH: 119036
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(119036)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-995-561-13314

  Query Match             4.8%;  Score 48.6;  DB 10;  Length 119036;
  Best Local Similarity   47.2%;  Pred. No. 12;
  Matches  244;  Conservative    1;  Mismatches  265;  Indels    7;  Gaps    3;

Qy        216 TTTTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGA 275
              ||| | ||  ||| |  | | |       |||||     || ||        || | | | |
Db      31804 TTTATATTTCAAATATATAAAATCTTATTTACATTTCATATATATATAAAATCTTATTTA 31745

Qy        276 TAATTCCTCTCTTTTTAAGAGGTTATTTTATA-ATTTTATAATAAAGCGCGGTTGTAGAA 334
              :| ||| | | | | ||| |  |||||||||| |||||    ||| |         | || ||
Db      31744 YATTTCATATATATATAAAATCTTATTTTATATATTTTTATATATATTTTATATATATAA 31685

Qy        335 AAGCATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAG 394
              || | |||   |||| ||   | |  |  ||||  | | | | |||  ||   |    |
Db      31684 AATCTTATTTTATATATATTTTATATATATTTTATATATATAAAATCATATTTTATATAT 31625

Qy        395 GTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCG 454
              |  || |     | | |||      |||||  |         |  |||||  ||
Db      31624 ATAAAATCTTATTTTATATATATTTTATATATTTTATATATAAAATCTTATTTTATATAT 31565

Qy        455 ATGTAAAACCATTGAAT---TTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAA 511
              || |  |   || ||     ||   |||| ||||||  ||| | |   | | |  |||
Db      31564 ATTTTTATATATAAAAAATCATATTTTATATATATTTTATATATTTTATATATATGAAAT 31505

Qy        512 CTCATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCATCAA 571
              || ||   |   | | |  ||||||  |   ||     | |    | ||       |  |
Db      31504 CTTATTTTATATATTATT---TTATATATATTTTATATATATATAAAATCTTATTTTATA 31448

Qy        572 TAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAATAATACAACATG 631
              || |  |  | | ||     |||| |  |   || ||   |  ||   ||| ||  |
Db      31447 TATATTTTATATATATATAAAATCTTATTTTATATATATTTTATATATATATAAAATCTT 31388

Qy        632 GTTTAAGTGAACGTGTAAAGTTCTTTGGGTTCGTGTCTGAAAAAGATTTACCATCGTTAT 691
              ||| |   |   | || |  | | |     || | | | | | | ||||   ||   |||
Db      31387 ATTTTATATATATTTTATATATATATAAAATCTTATTTTATATATATTTTATATATATAT 31328

Qy        692 ATAAGGGCTCGTTAGGATTAGTTTTCCCTTCTTTATA 728
              | ||      |||    || |||      | | || |
Db      31327 AAAATCTTATTTTATATATATTTTATATATATATAAA 31291
```

RESULT 4
US-10-301-480-568792
; Sequence 568792, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 568792
; LENGTH: 991
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-568792

Query Match 4.6%; Score 46.8; DB 12; Length 991;
Best Local Similarity 45.5%; Pred. No. 13;
Matches 214; Conservative 0; Mismatches 247; Indels 9; Gaps 1;

```
Qy        158 ATATTTTTCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTT 217
              ||||| |  |     |  |||  | || |  | ||   |  |   |  | |  | ||   |
Db        503 ATATTATATTAAATATAAAATATAATATTATATTATATATTATATTAAATATAACATAAT 562

Qy        218 TTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATA 277
                |  |     |||  |  ||   | |||| |  ||   ||| | |  | |  ||| |||
Db        563 AGTTTGTATTAAATATATTATATATAATATAATATAATAATGTTATATAATGTAATAATA 622

Qy        278 ATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAG 337
                |  | |  |  |   |  | ||| ||| ||||||||||| |       | || || |
Db        623 TATTATTTTATAATGTAACATAATTATATTATTTTATAATATATAATTTATTTAAAATAT 682

Qy        338 CATATAAAATAT---------TTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCAT 388
               ||||| |||||         | | |  |    || | |   |  |    |||
Db        683 AATATATAATATAATGTAATATAATATATGATTATATATTGTATTATGTTATATGTAATA 742

Qy        389 GGTCAGGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTA 448
                | |  | |||    | ||| | || | |          |||    ||| || | ||| ||
Db        743 TATTATATATAATATATGATATATTAATAATATATATAAATAATTTATATATTATATATA 802

Qy        449 ATGCCGATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAA 508
               ||    || ||  |   || ||  |    | || |||| |    | | ||  | | |  |
Db        803 ATTTATATATATTATATTTTAAAATGTAATTTATATATATAAATTATATATTATAAAATA 862

Qy        509 AAACTCATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCAT 568
               ||   |||      | ||          |   ||   ||    | || || ||  ||
Db        863 TAAAATATATAGTATATAATATATAATACATTATTTTGTTCATTTTAAATTTTTATATAT 922

Qy        569 CAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTA 618
               ||| |     | ||||      |||| | ||   || ||| | |  ||
Db        923 TTATATATTATATGTTTATATTGTATCCATATATAAATGTAGCATATATA 972
```

RESULT 5
US-10-301-480-568795
; Sequence 568795, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 568795
; LENGTH: 991
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-568795

Query Match 4.6%; Score 46.8; DB 12; Length 991;
Best Local Similarity 45.5%; Pred. No. 13;
Matches 214; Conservative 0; Mismatches 247; Indels 9; Gaps 1;

```
Qy        158 ATATTTTTCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTT 217
              ||||| |  |     |  |||  | || |  | ||   |  |   |  | |  | ||   |
Db        212 ATATTATATTAAATATAAAATATAATATTATATTATATATTATATTAAATATAACATAAT 271

Qy        218 TTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATA 277
                |  |     |||  |  ||   | |||| |  ||   ||| | |  | |  ||| |||
Db        272 AGTTTGTATTAAATATATTATATATAATATAATATAATAATGTTATATAATGTAATAATA 331

Qy        278 ATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAG 337
                |  | |  |  |   |  | ||| ||| ||||||||||| |       | || || |
Db        332 TATTATTTTATAATGTAACATAATTATATTATTTTATAATATATAATTTATTTAAAATAT 391

Qy        338 CATATAAAATAT---------TTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCAT 388
               ||||| |||||         | | |  |    || | |   |  |    |||
Db        392 AATATATAATATAATGTAATATAATATATGATTATATATTGTATTATGTTATATGTAATA 451

Qy        389 GGTCAGGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTA 448
                | |  | |||    | ||| | || | |          |||    ||| || | ||| ||
Db        452 TATTATATATAATATATGATATATTAATAATATATATAAATAATTTATATATTATATATA 511

Qy        449 ATGCCGATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAA 508
               ||    || ||  |   || ||  |    | || |||| |    | | ||  | | |  |
Db        512 ATTTATATATATTATATTTTAAAATGTAATTTATATATATAAATTATATATTATAAAATA 571

Qy        509 AAACTCATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCAT 568
               ||   |||      | ||          |   ||   ||    | || || ||  ||
Db        572 TAAAATATATAGTATATAATATATAATACATTATTTTGTTCATTTTAAATTTTTATATAT 631

Qy        569 CAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTA 618
               ||| |     | ||||      |||| | ||   || ||| | |  ||
Db        632 TTATATATTATATGTTTATATTGTATCCATATATAAATGTAGCATATATA 681
```

RESULT 6
US-10-301-480-1182201
; Sequence 1182201, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1182201
; LENGTH: 991
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-1182201

Query Match 4.6%; Score 46.8; DB 12; Length 991;

```
  Best Local Similarity   45.5%;  Pred. No. 13;
  Matches  214;  Conservative    0;  Mismatches  247;  Indels    9;  Gaps    1;

Qy        158 ATATTTTTCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTT 217
              ||||| | |     | ||| | || |  | ||   |  |   | | | | ||   |
Db        503 ATATTATATTAAATATAAAATATAATATTATATTATATATTATATTAAATATAACATAAT 562

Qy        218 TTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATA 277
              | |   ||| |  ||  | |||| |  ||  ||| | | | |  ||| |||
Db        563 AGTTTGTATTAAATATATTATATATAATATAATATAATAATGTTATATAATGTAATAATA 622

Qy        278 ATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAG 337
                | | |  |  |   |  | ||| ||| ||||||||||| |       | || || |
Db        623 TATTATTTTATAATGTAACATAATTATATTATTTTATAATATATAATTTATTTAAAATAT 682

Qy        338 CATATAAAATAT---------TTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCAT 388
              ||||| ||||          | | |  |    || | |    |  |     |||
Db        683 AATATATAATATAATGTAATATAATATATGATTATATATTGTATTATGTTATATGTAATA 742

Qy        389 GGTCAGGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTA 448
              | |  | |||    | ||| | || | |      |||   ||| || | ||| ||
Db        743 TATTATATATAATATATGATATATTAATAATATATATAAATAATTTATATATTATATATA 802

Qy        449 ATGCCGATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAA 508
              ||    || ||  |   || ||  |    | || |||| |    | | ||  | | |  |
Db        803 ATTTATATATATTATATTTTAAAATGTAATTTATATATATAAATTATATATTATAAAATA 862

Qy        509 AAACTCATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCAT 568
              ||   |||       | ||                |   ||   ||    | || || ||  ||
Db        863 TAAAATATATAGTATATAATATATAATACATTATTTTGTTCATTTTAAATTTTTATATAT 922

Qy        569 CAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTA 618
              |||  |    | ||||      |||| | ||   || ||| | |  ||
Db        923 TTATATATTATATGTTTATATTGTATCCATATATAAATGTAGCATATATA 972


RESULT 7
US-10-301-480-1182204
; Sequence 1182204, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1182204
;   LENGTH: 991
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1182204

  Query Match             4.6%;  Score 46.8;  DB 12;  Length 991;
  Best Local Similarity   45.5%;  Pred. No. 13;
  Matches  214;  Conservative    0;  Mismatches  247;  Indels    9;  Gaps    1;

Qy        158 ATATTTTTCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTATATTCCACCACTTT 217
              ||||| | |     | ||| | || |  | ||   |  |   | | | | ||   |
Db        212 ATATTATATTAAATATAAAATATAATATTATATTATATATTATATTAAATATAACATAAT 271

Qy        218 TTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCTGGATTTAAATGATA 277
              | |   ||| |  ||  | |||| |  ||  ||| | | | |  ||| |||
Db        272 AGTTTGTATTAAATATATTATATATAATATAATATAATAATGTTATATAATGTAATAATA 331

Qy        278 ATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAG 337
                | | |  |  |   |  | ||| ||| ||||||||||| |       | || || |
Db        332 TATTATTTTATAATGTAACATAATTATATTATTTTATAATATATAATTTATTTAAAATAT 391

Qy        338 CATATAAAATAT---------TTACAGTTTCGAATTTTTCAAAAGAAAGAATAGTAGCAT 388
              ||||| ||||          | | |  |    || | |    |  |     |||
Db        392 AATATATAATATAATGTAATATAATATATGATTATATATTGTATTATGTTATATGTAATA 451

Qy        389 GGTCAGGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCTATTTA 448
              | |  | |||    | ||| | || | |      |||   ||| || | ||| ||
Db        452 TATTATATATAATATATGATATATTAATAATATATATAAATAATTTATATATTATATATA 511

Qy        449 ATGCCGATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAACAGAA 508
              ||    || ||  |   || ||  |    | || |||| |    | | ||  | | |  |
Db        512 ATTTATATATATTATATTTTAAAATGTAATTTATATATATAAATTATATATTATAAAATA 571

Qy        509 AAACTCATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGATCCAT 568
              ||   |||       | ||                |   ||   ||    | || || ||  ||
Db        572 TAAAATATATAGTATATAATATATAATACATTATTTTGTTCATTTTAAATTTTTATATAT 631

Qy        569 CAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTA 618
              |||  |    | ||||      |||| | ||   || ||| | |  ||
Db        632 TTATATATTATATGTTTATATTGTATCCATATATAAATGTAGCATATATA 681


RESULT 8
US-09-925-065A-201658/c
; Sequence 201658, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 201658
;   LENGTH: 325
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-201658

  Query Match             4.5%;  Score 46;  DB 7;  Length 325;
  Best Local Similarity   49.2%;  Pred. No. 16;
  Matches  121;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy        295 AGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACA 354
              | | |  |||| |  ||    ||||| |  |||   || ||    | | || | | |
Db        291 AAGCTTATTTAAAGGTTATGTATAAAACAAGGTAAAAGGAACCAGGAAATAAGAGAGATA 232

Qy        355 GTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATA 414
              || || |  ||| |||||     |   |||   |||   ||  || | | | |  |
Db        231 TAAAGAAAGTTATAGAAATAAAGAGGGTTTTTTTGGGGGGGTAAGAAAGCTTAAAGAGAA 172

Qy        415 GTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTA 474
              | |   |||||||     | |||| || |  | |||   ||  || ||  |    | |
Db        171 ATAATTTTATATAAGAAAGAATCTTGTATGGTAAATTTAGACCTAGAATAAAATGACTAG 112

Qy        475 GGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTT 534
              | |||  |   |  |||| | ||| | |  | ||      ||| |||  | |  || |
Db        111 TTGTTTAAKAAAGAGGAATGTTTGGGACAAACCAGAAAGTCCAAGCATGTCATGAATGGT 52

Qy        535 ATATCT 540
```

```
                | | |
Db          51 CTGTGT 46
```

RESULT 9
US-10-301-480-290170/c
; Sequence 290170, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 290170
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-290170

Query Match 4.5%; Score 46; DB 12; Length 336;
Best Local Similarity 49.2%; Pred. No. 16;
Matches 121; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```
Qy        295 AGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACA 354
              | | |  |||| |  ||    |||||  |  |||   || ||     | | || |   | |
Db        302 AAGCTTATTTAAAGGTTATGTATAAAACAAGGTAAAAGGAACCAGGAAATAAGAGAGATA 243

Qy        355 GTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATA 414
                || || |  ||| |||||     |     |||   |||   ||  || | | |  |
Db        242 TAAAGAAAGTTATAGAAATAAAGAGGGTTTTTTTGGGGGGGTAAGAAAGCTTAAAGAGAA 183

Qy        415 GTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTA 474
              | |   |||||||    | |||| || |  | |||   ||  || ||  |    | |
Db        182 ATAATTTTATATAAGAAAGAATCTTGTATGGTAAATTTAGACCTAGAATAAAATGACTAG 123

Qy        475 GGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTT 534
                | |||  |   |  |||| | ||| | |  | ||     ||| |||  | |  || |
Db        122 TTGTTTAAKAAAGAGGAATGTTTGGGACAAACCAGAAAGTCCAAGCATGTCATGAATGGT 63

Qy        535 ATATCT 540
                | | |
Db         62 CTGTGT 57
```

RESULT 10
US-10-301-480-903579/c
; Sequence 903579, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 903579
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-903579

Query Match 4.5%; Score 46; DB 12; Length 336;
Best Local Similarity 49.2%; Pred. No. 16;
Matches 121; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```
Qy        295 AGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACA 354
              | | |  |||| |  ||    |||||  |  |||   || ||     | | || |   | |
Db        302 AAGCTTATTTAAAGGTTATGTATAAAACAAGGTAAAAGGAACCAGGAAATAAGAGAGATA 243

Qy        355 GTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATA 414
                || || |  ||| |||||     |     |||   |||   ||  || | | |  |
Db        242 TAAAGAAAGTTATAGAAATAAAGAGGGTTTTTTTGGGGGGGTAAGAAAGCTTAAAGAGAA 183

Qy        415 GTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTA 474
              | |   |||||||    | |||| || |  | |||   ||  || ||  |    | |
Db        182 ATAATTTTATATAAGAAAGAATCTTGTATGGTAAATTTAGACCTAGAATAAAATGACTAG 123

Qy        475 GGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTT 534
                | |||  |   |  |||| | ||| | |  | ||     ||| |||  | |  || |
Db        122 TTGTTTAAKAAAGAGGAATGTTTGGGACAAACCAGAAAGTCCAAGCATGTCATGAATGGT 63

Qy        535 ATATCT 540
                | | |
Db         62 CTGTGT 57
```

RESULT 11
US-11-011-332A-95
; Sequence 95, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
; APPLICANT: Foekens, John
; APPLICANT: Harbeck, Nadia
; APPLICANT: Koenig, Thomas
; APPLICANT: Maier, Sabine
; APPLICANT: Martens, John
; APPLICANT: Model, Fabian
; APPLICANT: Nimmrich, Inko
; APPLICANT: Rujan, Tamas
; APPLICANT: Schmitt, Manfred
; APPLICANT: Lesche, Ralf
; APPLICANT: Dietrich, Dimo
; APPLICANT: Mueller, Volkmar
; APPLICANT: Kluth, Antje
; APPLICANT: Schwope, Ina
; APPLICANT: Hartmann, Oliver
; APPLICANT: Adorjan, Peter
; TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SUF
; TITLE OF INVENTION: BREAST CELL PROLIFERATIVE DISORDER PATIENTS
; FILE REFERENCE: 47675-99
; CURRENT APPLICATION NUMBER: US/11/011,332A
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 10/517,741
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: PCT/EP2003/010881
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: DE 10245779.4
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: DE 10300096.8
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: DE 10317955.0
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: PCT/EP2004/014170
; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: EP 03090432.0
; PRIOR FILING DATE: 2003-12-11
; PRIOR APPLICATION NUMBER: EP 04090041.7
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: EP 04090380.9
; PRIOR FILING DATE: 2004-09-30
; PRIOR APPLICATION NUMBER: EP 04090127.4
; PRIOR FILING DATE: 2004-04-01

```
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 95
;   LENGTH: 6257
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-95

  Query Match             4.5%;  Score 46;  DB 17;  Length 6257;
  Best Local Similarity   45.0%;  Pred. No. 24;
  Matches  214;  Conservative    0;  Mismatches  260;  Indels    2;  Gaps    1;

Qy        144 AGGTTTAGCAGTCAATATTTTTCTGCGAAAGAAAACTGATGTGGTTTTTCTTCCTGGCTA 203
              ||    | |  |  ||  || ||      || || |    |||  ||||| ||  | | |
Db          2 AGAAGTGGATGGAAAGGTTATTTGTTTGAATAAGAGGAGTGTTTTTTTTTTTATTTGGTT 61

Qy        204 TATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATAACAATACATGATCTAAATCATCT 263
              ||      |||   |||      | || || ||            ||   | |     |
Db         62 ATTTGTTTGTTTTTAAGGTTTTGTTTAATTTATTATTGTTGGTTGGAATAAGATTTTAAT 121

Qy        264 GGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTATTTTATAATTTTATAATAAAGCG 323
              |||  |   |  ||  | |    || | |  || |||    ||| |     ||    |
Db        122 TATTTTTTAATGTTTTTTTTTTAGAGTTTAAAAATTTTTTGGATATTATTGGGTAGTTGG 181

Qy        324 CGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCGAATTTTTCAAAAGAAAGAATAGT 383
              |||  | ||  |||    |  | | |||  | |||  | |||||  |    |      || |
Db        182 GAGTTTTGGATAAGTTGTTTTATTTTTTTGAATTTTTATTTTTTTATTTGTGTATTTATT 241

Qy        384 AGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACGGTATATAATGGGGTATCTAGTCT 443
                ||   | | ||  |   ||  ||    |  ||   | |||   |||| |||
Db        242 TTATAGGATTGTTATAGAGATTA--AATGAAAGTATGGATAAAAATTTAGTATATAGGAA 299

Qy        444 ATTTAATGCCGATGTAAAACCATTGAATTTAGGCTATAAATATTTGCTATGTGTAGGAAA 503
              | || ||   ||  |    |  | | ||  || | ||     ||| |  || |  || |
Db        300 GTATATTGTATATTTTTTAAAAGTTAAGATAAGATAGGTTTATATAGTAGGAATATGGGT 359

Qy        504 CAGAAAAACTCATAAGAATGAGAAGTGTGTTATATCTGCCTTTGCCAAAGCAGATATTGA 563
              |    || | || ||||    || | |||| | | |    ||    ||     || |
Db        360 AATTTTAAATTATGAGAAAATTAAATATGTTTTGTGTTTGATTTTGGAATGTTATGTATT 419

Qy        564 TCCATCAATAAAACTCGTTTTTACTGGTAATCCTTGTAATGATTTAGAAAAACTAA 619
              |  |  | | | | | ||||  |  |  |  |||| |  ||| || |||  ||
Db        420 TTTTTTTTTATATTTGATTTTTATTTTTTTAATTGTTTTTTTTTTGATAAATAAA 475


RESULT 12
US-09-925-065A-165636/c
; Sequence 165636, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 165636
;   LENGTH: 420
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-165636

  Query Match             4.5%;  Score 45.6;  DB 7;  Length 420;
  Best Local Similarity   48.8%;  Pred. No. 19;
  Matches  120;  Conservative    1;  Mismatches  125;  Indels    0;  Gaps    0;

Qy        295 AGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACA 354
              | | |  |||| |  ||     ||||| |  |||   || ||     | | || |   | |
Db        318 AAGCTTATTTAAAGGTTATGTATAAAACAAGGTAAAAGGAACCAGGAAATAAGAGAGATA 259

Qy        355 GTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATA 414
                || || |  ||| |||||     |    |:|   |||   ||  || | | |  |
Db        258 TAAAGAAAGTTATAGAAATAAAGAGGGTTTTTTTKGGGGGGTAAGAAAGCTTAAAGAGAA 199

Qy        415 GTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTA 474
              | |   |||||||    | |||| || |  | |||   ||  || ||  |    | |
Db        198 ATAATTTTATATAAGAAAGAATCTTGTATGGTAAATTTAGACCTAGAATAAAATGACTAG 139

Qy        475 GGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTT 534
              | |||  |   |  |||| | ||| | |  | ||      ||| |||  | |  || |
Db        138 TTGTTTAAGAAAGAGGAATGTTTGGGACAAACCAGAAAGTCCAAGCATGTCATGAATGGT 79

Qy        535 ATATCT 540
              | | |
Db         78 CTGTGT 73


RESULT 13
US-10-301-480-257920/c
; Sequence 257920, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 257920
;   LENGTH: 427
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-257920

  Query Match             4.5%;  Score 45.6;  DB 12;  Length 427;
  Best Local Similarity   48.8%;  Pred. No. 19;
  Matches  120;  Conservative    1;  Mismatches  125;  Indels    0;  Gaps    0;

Qy        295 AGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACA 354
              | | |  |||| |  ||     ||||| |  |||   || ||     | | || |   | |
Db        318 AAGCTTATTTAAAGGTTATGTATAAAACAAGGTAAAAGGAACCAGGAAATAAGAGAGATA 259

Qy        355 GTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATA 414
                || || |  ||| |||||     |    |:|   |||   ||  || | | |  |
Db        258 TAAAGAAAGTTATAGAAATAAAGAGGGTTTTTTTKGGGGGGTAAGAAAGCTTAAAGAGAA 199

Qy        415 GTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTA 474
              | |   |||||||    | |||| || |  | |||   ||  || ||  |    | |
Db        198 ATAATTTTATATAAGAAAGAATCTTGTATGGTAAATTTAGACCTAGAATAAAATGACTAG 139

Qy        475 GGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTT 534
              | |||  |   |  |||| | ||| | |  | ||      ||| |||  | |  || |
Db        138 TTGTTTAAGAAAGAGGAATGTTTGGGACAAACCAGAAAGTCCAAGCATGTCATGAATGGT 79

Qy        535 ATATCT 540
              | | |
```

```
Db         78 CTGTGT 73


RESULT 14
US-10-301-480-871329/c
; Sequence 871329, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 871329
;   LENGTH: 427
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-871329

  Query Match             4.5%;  Score 45.6;  DB 12;  Length 427;
  Best Local Similarity   48.8%;  Pred. No. 19;
  Matches  120;  Conservative    1;  Mismatches  125;  Indels    0;  Gaps    0;

Qy        295 AGGTTATTTTATAATTTTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACA 354
              | | |  |||| |  ||    ||||| |  |||   || ||      | | || |   | |
Db        318 AAGCTTATTTAAAGGTTATGTATAAAACAAGGTAAAAGGAACCAGGAAATAAGAGAGATA 259

Qy        355 GTTTCGAATTTTTCAAAAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATA 414
                || ||  |  ||| |||||     |     |:|   |||   ||  || | | |  |
Db        258 TAAAGAAAGTTATAGAAATAAAGAGGGTTTTTTTKGGGGGGTAAGAAAGCTTAAAGAGAA 199

Qy        415 GTCACGGTATATAATGGGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTA 474
               | |   |||||||    | |||| || |  | |||   ||  || ||  |    | |
Db        198 ATAATTTTATATAAGAAAGAATCTTGTATGGTAAATTTAGACCTAGAATAAAATGACTAG 139

Qy        475 GGCTATAAATATTTGCTATGTGTAGGAAACAGAAAAACTCATAAGAATGAGAAGTGTGTT 534
              | |||  |   |  |||| | ||| | |  | ||     ||| |||  | |  || |
Db        138 TTGTTTAAGAAAGAGGAATGTTTGGGACAAACCAGAAAGTCCAAGCATGTCATGAATGGT 79

Qy        535 ATATCT 540
                | | |
Db         78 CTGTGT 73


RESULT 15
US-10-301-480-568793
; Sequence 568793, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 568793
;   LENGTH: 990
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-568793

  Query Match             4.5%;  Score 45;  DB 12;  Length 990;
  Best Local Similarity   45.6%;  Pred. No. 29;
  Matches  194;  Conservative    0;  Mismatches  230;  Indels    1;  Gaps    1;

Qy         72 AATTTCTAAACTACTATGTGATATATCTAAGGAGGAACTATATAGAAAATGTGCTTCTCC 131
              ||| |  ||  || ||| |  |||||  ||      ||  ||||| || ||     |
Db        299 AATATAAAATATAATATATATTATATTAAATATAAAATAATATAAAATATAATATATATT 358

Qy        132 GCTGGCCCCATTAGGTTTAGCAGTCAATAT-TTTTCTGCGAAAGAAAACTGATGTGGTTT 190
                  || |  | ||  |   | ||| |||| |   ||| | ||   |||| | |
Db        359 ATATTAAATATAAAATATAATATATATTATATTTTATATTAAATATAAAATATGTTATAT 418

Qy        191 TTCTTCCTGGCTATATTCCACCACTTTTTTGTTCGAAAAAGTTCATAATAACAATACATG 250
               |  |  |   ||||    |  |  || |  |    |  |  | | |   |  |||   |
Db        419 ATTATATTAAATATAAAATATAATATTATATTATATATTATATTAAATATAACATAATAG 478

Qy        251 ATCTAAATCATCTGGATTTAAATGATAATTCCTCTCTTTTTAAGAGGTTATTTTATAATT 310
               |   | | |        | | || |||          | | |||     || | ||| |||
Db        479 TTTGTATTAAATATATATAATATAATATAATAATGTTATATAATGTAATAATATATTATT 538

Qy        311 TTATAATAAAGCGCGGTTGTAGAAAAGCATATAAAATATTTACAGTTTCGAATTTTTCAA 370
              |||||||  | |    || ||  |    |||  | ||| ||    |     ||  |  |
Db        539 TTATAATGTAACATAATTATATTATTTTATAATATATAATTTATTTAAAATATAATATAT 598

Qy        371 AAGAAAGAATAGTAGCATGGTCAGGTGTAAACCCTAATAAAATAGTCACGGTATATAATG 430
              || | |   || ||  ||       | || |   || ||   ||       ||||| ||
Db        599 AATATAATGTAATATAATATATGATTATATATTGTATTATGTTATATGTAATATATTATA 658

Qy        431 GGGTATCTAGTCTATTTAATGCCGATGTAAAACCATTGAATTTAGGCTATAAATATTTGC 490
               || ||   ||| | |     || || ||  |   || ||   |||| |  ||
Db        659 TATAATATATGATATATTAATAATATATATAAATAATTTATATATTATATATAATTTATA 718

Qy        491 TATGT 495
                |||
Db        719 TATAT 723


Search completed: May 12, 2006, 06:55:01
Job time : 518.104 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 08:31:04 ; Search time 4168.05 Seconds
(without alignments)
15792.702 Million cell updates/sec

Title: US-09-423-093B-3_COPY_7310_8467
Perfect score: 1158
Sequence: 1 atggaaataaatgaaataaa.....aacagattattaaacgttaa 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1158	100.0	12441	1	SERFBC2	X61917 S.enterica
2	1158	100.0	12441	6	BD085556	BD085556 Nucleic a
3	1158	100.0	12446	6	AX029263	AX029263 Sequence
4	556	48.0	16332	1	AF104912	AF104912 Escherich
5	520	44.9	1167	6	AR384270	AR384270 Sequence
6	398	34.4	21411	1	AB041266	AB041266 Actinobac
7	340.6	29.4	110000	1	AE016853_38	Continuation (39 o
8	340.2	29.4	110000	1	CP000075_38	Continuation (39 o
9	332.6	28.7	110000	1	CP000058_36	Continuation (37 o
10	269	23.2	8375	1	AE009256	AE009256 Agrobacte
c 11	269	23.2	11355	1	AE008357	AE008357 Agrobacte
c 12	267.6	23.1	110000	1	BX571966_23	Continuation (24 o
c 13	267.6	23.1	110000	1	CP000011_18	Continuation (19 o
14	257.4	22.2	110000	1	CP000010_00	CP000010 Burkholde
15	257.2	22.2	110000	1	BX571965_06	Continuation (7 of
c 16	253.2	21.9	12927	1	AE007804	AE007804 Clostridi
c 17	231.2	20.0	110000	1	BA000019_57	Continuation (58 o
c 18	227.2	19.6	110000	1	BA000022_20	Continuation (21 o
19	224.6	19.4	15890	1	AB046360	AB046360 Actinobac
c 20	219.4	18.9	110000	1	CP000095_08	Continuation (9 of
c 21	218.2	18.8	300143	1	AE017165	AE017165 Prochloro
22	183.2	15.8	110000	1	RME591985_07	Continuation (8 of
c 23	181.4	15.7	324	6	AR384476	AR384476 Sequence
24	178.8	15.4	349323	1	BX572096	BX572096 Prochloro
25	173.4	15.0	304923	1	AE017289	AE017289 Leptospir
c 26	171.8	14.8	10582	1	AE011513	AE011513 Leptospir
c 27	160	13.8	349213	1	BX569693	BX569693 Synechoco
28	150	13.0	110000	1	BA000019_34	Continuation (35 o
c 29	146.4	12.6	110000	1	AP008231_22	Continuation (23 o
30	139.4	12.0	213732	1	AE001862	AE001862 Deinococc
c 31	120.2	10.4	222	6	AR384397	AR384397 Sequence
32	105.2	9.1	110000	1	BA000045_16	Continuation (17 o
c 33	101.8	8.8	295650	1	BX294151	BX294151 Pirellula
34	101	8.7	349635	1	BX572603	BX572603 Rhodopseu
35	90.6	7.8	2042	1	U75690	U75690 Nostoc sp.
36	75.4	6.5	15564	1	AB185495	AB185495 Corynebac
37	59.8	5.2	1141	6	AR579680	AR579680 Sequence
38	59.8	5.2	1141	6	AX083744	AX083744 Sequence
39	59.2	5.1	13910	1	AE001082	AE001082 Archaeogl
40	58.4	5.0	110000	1	CP000099_04	Continuation (5 of
41	55.2	4.8	11810	1	AE010785	AE010785 Methanosa
c 42	55.2	4.8	300343	1	AE017310	AE017310 Desulfovi
43	53	4.6	80106	8	AL645634	AL645634 Human DNA
44	50.2	4.3	162935	5	CR847843	CR847843 Zebrafish
45	48	4.1	110000	1	AP006716_05	Continuation (6 of

ALIGNMENTS

RESULT 1
SERFBC2
LOCUS SERFBC2 12441 bp DNA linear BCT 09-SEP-2004
DEFINITION S.enterica rfbJ gene cluster.
ACCESSION X61917
VERSION X61917.1 GI:47004
KEYWORDS abe gene; abequose synthase; galactosyltransferase; manB gene; manC gene; mannose-1-phosphate guanyltransferase; phosphomannomutase; wbaL gene; wbaP gene; wbaQ gene; wbaR gene; wbaW gene; wbaZ gene; wzx gene; wzy gene.
SOURCE Salmonella enterica
ORGANISM Salmonella enterica
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 12441)
AUTHORS Brown,P.K., Romana,L.K. and Reeves,P.R.
TITLE Molecular analysis of the rfb gene cluster of Salmonella serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B
JOURNAL Mol. Microbiol. 6 (10), 1385-1394 (1992)
PUBMED 1379320
REFERENCE 2
AUTHORS Reeves,P.
TITLE Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale
JOURNAL Trends Genet. 9 (1), 17-22 (1993)
PUBMED 8434412
REFERENCE 3
AUTHORS Liu,D., Haase,A.M., Lindqvist,L., Lindberg,A.A. and Reeves,P.R.
TITLE Glycosyl transferases of O-antigen biosynthesis in Salmonella enterica: identification and characterization of transferase genes of groups B, C2, and E1
JOURNAL J. Bacteriol. 175 (11), 3408-3413 (1993)
PUBMED 7684736
REFERENCE 4 (bases 1 to 12441)
AUTHORS Brown,P.K.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1991) P.K. Brown, University of Sydney, Dept of Microbiology, Maze Crescent, Sydney N S W 2006, AUSTRALIA
FEATURES Location/Qualifiers
source 1. .12441

/organism="Salmonella enterica"
/mol_type="genomic DNA"
/strain="group C2, serovar muenchen M67"
/db_xref="taxon:28901"
gene 138. .1019
/gene="abe"
CDS 138. .1019
/gene="abe"
/codon_start=1
/transl_table=11
/product="abequose synthase"
/protein_id="CAA43918.1"
/db_xref="GI:47005"
/db_xref="GOA:Q00329"
/db_xref="UniProt/Swiss-Prot:Q00329"
/translation="MLDVNKKILMTGATSFVGTHLLHSLIKEGYSIIALKRPITEPTI
INTLIEWLNIQDIEKICQSSMNIHAIVHIATDYGRNRTPISEQYKCNVLLPTRLLELM
PALKTKFFISTDSFFGKYEKHYGYMRSYMASKRHFVELSKIYVEEHPDVCFINLRLEH
VYGERDKAGKIIPYVIKKMKNNEDIDCTIARQKRDFIYIDDVVSAYLKILKEGFNAGH
YDVEVGTGKSIELKEVFEIIKKETHSSSKINYGAVAMRDDEIMESHANTSFLTRLGWS
AEFSIEKGVKKMLSMKE"
CDS 1019. .2359
/note="wzx
Referred to as orf12.6 in reference [2]"
/citation=[2]
/codon_start=1
/transl_table=11
/product="hypothetical protein with 12 predicted
transmembrane segments which possibly functions in export
of O-antigen"
/protein_id="CAA43919.1"
/db_xref="GI:47006"
/db_xref="UniProt/TrEMBL:Q00339"
/translation="MNRIIRMLGVDKAIRYVIFGKIISVLTGLLLIMLISHHLSKDAQ
GYYYTFNSVVALQIIFELGLSTVIIQFASHEMSALKYDYSERDIIGESKNKQRYLSLF
RLAIKWYAVIALLIILIVGPIGYVFFTQKEGLGVPWQGAWLLLTIVTAFNIFLVSVLS
VAEGSGLITDVNKMRMYQSLLAGILAVSLLISGFGLYATSAIAISGTIIFSIFSYKYF
KKIFLQSLKHKNKYTEGGISWVNEIFPMQWRIALSWMSGYFIYFVMTPIAFKYFGAIY
AGQLGMSLTLCNMVMATGLAWISTKYPKWGVMVSNKQLAELSKSFKSAVMQSSFFVLT
GLTGVYISLWLLKLSGSNIGERFLGLQDFFFLSLAIIGNHIVACFATYIRAHKTEKMT
LASCIMALLTITTMLFVAYLEYSRFYMLMYAALTWLYFVPQTYIIFKRFKSSYE"
CDS 2352. .3314
/note="unnamed protein product; wbaR"
/codon_start=1
/transl_table=11
/protein_id="CAA43909.1"
/db_xref="GI:47007"
/db_xref="UniProt/TrEMBL:Q00337"
/translation="MSKKPLLTIAIPTYNRSSCLARLLDSIIQQENYCHDELEVIVCD
NASTDETARIAKSGLDKIRNSTYHLNEENLGMDGNFQKCFELSNGKYLWMIGDDDLIV
KNGISKVFSILKSRPALDMVYVNSAAKTELNYNADVRTSFYTNDVDFISDVKVMFTFI
SGMICKKTDAIVKAVGIFSPQTTGKYLMHLTWQLPLLKQGGEFAVIHNNIIEAEPDNS
GGYHLYKVFSNNLATIFDVFYPREHRVSKRVRASACLFLLNFIGDEDKTKNFATNNYL
RDCDSAFIDLIIYKYGLRFFYLYPKTVPLFRKIKYIIKTVLMRK"
CDS 3361. .3975
/note="unnamed protein product; wbaL
orf14.9 possibly encodes an acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="CAA43910.1"
/db_xref="GI:47008"
/db_xref="UniProt/TrEMBL:Q00480"
/translation="MFVYSLRLKLNLIISLLSKVRRKSKAKFLVLLSGYDFKMVGKNF
KLNVKPYSAKNNTSSKWGSMRVGDNCWIEAVYNYGDEKFEPYLYIGDRICLSDNVHIS
CVSCLILENDILIGSKVYIGDHSHGSYKVCSPKIEPPANKPLGDIAPIKIGNCCWIGD
NAVILAGSEICDGCVIAANSVVKDLKVDKPCLIGGVPAKVIKVF"
CDS 3977. .5020
/note="wbaQ
Referred to as orf15.6 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="rhamnosyl transferase"
/protein_id="CAA43911.1"
/db_xref="GI:47009"
/db_xref="GOA:Q00482"
/db_xref="UniProt/TrEMBL:Q00482"
/translation="MNVFISICIPSYNRAEFLEPLLDSIYNQDYCLKNNDFEVIVCED
KSPQRDEINSIIENYKAKNNKQNLYVNFNEDNLGYDKNLKKCISLTTGKYCMIMGNDD
LLADGALSKIVKVLKANPEIVLATRAYGWFKENPNELCDTVRHLTDDTLFQPGADAIK
FFFRRVGVISGFIVNAEKAKKLSSDLFDGRLYYQMYLAGMLMAEGQGYYFSDVMTLSR
DTEAPDFGNAGTEKGVFTPGGYKPEGRIHMVEGLLLIAKYIEDTTKIDGVYAGIRKDL
ANYFYPYIRDQLDLPLYTYIKMINKFRKMGFSNEKLFYVHAFLGYVLKRRGYDALIKY
IRSKKGGTPRLGI"
gene 5114. .6313
/gene="wzy"
CDS 5114. .6313
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O antigen polymerase"
/protein_id="CAA43912.1"
/db_xref="GI:47010"
/db_xref="GOA:Q00474"
/db_xref="UniProt/Swiss-Prot:Q00474"
/translation="MLPFPPGAILRDVLNVFFVALVLVRFVIDRKKTYFPLVFTIFSW
SAVILWVIALTIFSPDKIQAIMGGRSYILFPAVFIALVILKVSYPQSLNIEKIVCYII
FLMFMVATISIIDVLMNGEFIKLLGYDEHYAGEQLNLINSYDGMVRATGGFSDALNFG
YMLTLGVLLCMECFSQGYKRLLMLIISFVLFIAICMSLTRGAILVAALIYALYIISNR
KMLFCGITLFVIIPVLAISTNIFDNYTEILIGRFTDSSQASRGSTQGRIDMAINSLN
FLSEHPSGIGLGTQGSGNMLSVKDNRLNTDNYFFWIALETGIIGLIINIIYLASQFYS
STLLNRIYGSHCSNMHYRLYFLFGSIYFISAALSSAPSSSTFSIYYWTVLALIPFLKL
TNRRCTR"
CDS 6313. .7323
/note="wbaW
Referred to as orf17.9 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="second mannosyl transferase"
/protein_id="CAA43913.1"
/db_xref="GI:47011"
/db_xref="GOA:Q00481"
/db_xref="UniProt/TrEMBL:Q00481"
/translation="MNNKKVLMDISWSNKGGIGRFTDEISKLLCDISKEELYRKCASP
LAPLGLAVNIFLRKKTDVVFLPGYIPPLFCSKKFIITIHDLNHLDLNDNSSLFKRLFY
NFIIKRGCRKAYKIFTVSNFSKERIVAWSGVNPNKIVTVYNGVSSLFNADVKPLNLGY
KYLLCVGNRKTHKNEKCVISAFAKADIDPSIKLVFTGNPCNDLEKLIIQHGLSERVKF
FGFVSEKDLPSLYKGSLGLVFPSLYEGFGLPVVEGMACGIPVLTSLTSSLPEVAGDAA
ILVDPLSEDAITKGISRLINDSELRKHLIQKGLLRAKRFNWQNVVSEIEMVLTEACDG
NK"
CDS 7310. .8467
/note="wbaZ
Referred to as orf18.9 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="first mannosyl transferase"
/protein_id="CAA43914.1"
/db_xref="GI:47012"
/db_xref="GOA:Q00483"
/db_xref="UniProt/TrEMBL:Q00483"
/translation="MEINEIKISLVHEWLLSYAGSEQVSSAILHVFPEAKLYSVVDFL
TDEQRRHFLGKYATTTFIQNLPKAKKFYQKYLPLMPLAIEQLDLSDANIIISSAHSVA
KGVISGPDQLHISYVHSPIRYAWDLQHQYLNESNLNKGIKGWLAKWLLHKIRIWDSRT
ANGVDHFIANSQYIARRIKKVYRREASVIYPPVDVDNFEVKNEKQDYYFTASRMVPYK
RIDLIVEAFSKMPEKKLVVIGDGPEMKKIKSKATDNIKLLGYQSFPVLKEYMQSARAF
VFAAEEDFGIIPVEAQACGTPVIAFGKGGALETVRPLGVEEPTGIFFKEQNIASLHEA
VSEFEKNASFFTSQACRKNAEKFSRSRFEQEFKNFVNEKWNLFKTEQIIKR"
gene 8482. .9903
/gene="manC"
CDS 8482. .9903
/gene="manC"
/codon_start=1
/transl_table=11
/product="mannose-1-phosphate guanyltransferase"

```
                     /protein_id="CAA43915.1"
                     /db_xref="GI:47013"
                     /db_xref="GOA:Q00473"
                     /db_xref="UniProt/Swiss-Prot:Q00473"
                     /translation="MSKLIPVIMAGGIGSRLWPLSREEHPKQFLSVDGELSMLQNTIK
                     RLTPLLAGEPLVICNDSHRFLVAEQLRAINKLANNIILEPVGRNTAPAIALAAFCSLQ
                     NVVDEDPLLLVLAADHVIRDEKVFLKAINHAEFFATQGKLVTFGIVPTQAETGYGYIC
                     RGEAIGEDAFSVAEFVEKPDFDTARHYVESEKYYWNSGMFLFRASSYLQELKDLSPDI
                     YQACENAVGSINPDLDFIRIDKEAFAMCPSDSIDYAVMEHTRHAVVVPMNAGWSDVGS
                     WSSLWDISKKDPQRNVLHGDIFAYNSKDNYIYSEKSFISTIGVNNLVIVQTADALLVS
                     DKDSVQDVKKVVDYLKANNRNEHKKHLEVFRPWGKFSVIHSGDNYLVKRITVKPGAKF
                     AAQMHLHRAEHWIVVSGTACITKGEEIFTISENESTFIPANTVHTLKNPATIPLELIE
                     IQSGTYLAEDDIIRLEKHSGYLE"
     gene            9911. .11347
                     /gene="manB"

  Query Match           100.0%;  Score 1158;  DB 1;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 9.1e-219;
  Matches 1158;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7310 ATGGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGC 7369

Qy         61 TCCGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7370 TCCGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTG 7429

Qy        121 GTTGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7430 GTTGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACA 7489

Qy        181 TTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7490 TTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCA 7549

Qy        241 CTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7550 CTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCC 7609

Qy        301 GTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCT 7669

Qy        361 ATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 ATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGA 7729

Qy        421 ATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 ATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACC 7789

Qy        481 GCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAA 7849

Qy        541 GTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 GTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTA 7909

Qy        601 AAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 AAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATT 7969

Qy        661 GATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGAT 8029

Qy        721 GGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTAT 8089

Qy        781 CAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCG 8149

Qy        841 GAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCC 8209

Qy        901 TTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGC 8269

Qy        961 ATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 ATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAA 8329

Qy       1021 AATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 AATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCA 8389

Qy       1081 AGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAA 8449

Qy       1141 CAGATTATTAAACGTTAA 1158
              ||||||||||||||||||
Db       8450 CAGATTATTAAACGTTAA 8467


RESULT 2
BD085556
LOCUS       BD085556               12441 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085556
VERSION     BD085556.1  GI:22631166
KEYWORDS    JP 2001522248-A/3.
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 12441)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 3 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Salmonella enterica
            PN   JP 2001522248-A/3
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .12441
            FT                   /organism='Salmonella enterica'.
FEATURES             Location/Qualifiers
     source          1. .12441
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match           100.0%;  Score 1158;  DB 6;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 9.1e-219;
  Matches 1158;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7310 ATGGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGC 7369

Qy         61 TCCGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTG 120
```

```
Db       7370 TCCGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTG 7429

Qy        121 GTTGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7430 GTTGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACA 7489

Qy        181 TTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7490 TTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCA 7549

Qy        241 CTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7550 CTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCC 7609

Qy        301 GTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCT 7669

Qy        361 ATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 ATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGA 7729

Qy        421 ATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 ATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACC 7789

Qy        481 GCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAA 7849

Qy        541 GTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 GTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTA 7909

Qy        601 AAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 AAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATT 7969

Qy        661 GATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGAT 8029

Qy        721 GGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTAT 8089

Qy        781 CAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCG 8149

Qy        841 GAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCC 8209

Qy        901 TTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGC 8269

Qy        961 ATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 ATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAA 8329

Qy       1021 AATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 AATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCA 8389

Qy       1081 AGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAA 8449

Qy       1141 CAGATTATTAAACGTTAA 1158
              ||||||||||||||||||

Db       8450 CAGATTATTAAACGTTAA 8467


RESULT 3
AX029263
LOCUS       AX029263               12446 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 3 from Patent WO9850531.
ACCESSION   AX029263
VERSION     AX029263.1  GI:10190114
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 3 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .12446
                     /organism="Salmonella enterica"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match            100.0%;  Score 1158;  DB 6;  Length 12446;
  Best Local Similarity  100.0%;  Pred. No. 9.1e-219;
  Matches 1158;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7310 ATGGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGC 7369

Qy         61 TCCGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7370 TCCGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTG 7429

Qy        121 GTTGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7430 GTTGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACA 7489

Qy        181 TTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7490 TTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCA 7549

Qy        241 CTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7550 CTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCC 7609

Qy        301 GTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCT 7669

Qy        361 ATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 ATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGA 7729

Qy        421 ATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 ATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACC 7789

Qy        481 GCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAA 7849

Qy        541 GTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 GTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTA 7909

Qy        601 AAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       7910 AAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATT 7969

Qy        661 GATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGAT 8029

Qy        721 GGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTAT 8089

Qy        781 CAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCG 8149

Qy        841 GAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCC 8209

Qy        901 TTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGC 8269

Qy        961 ATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 ATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAA 8329

Qy       1021 AATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 AATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCA 8389

Qy       1081 AGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAA 8449

Qy       1141 CAGATTATTAAACGTTAA 1158
              ||||||||||||||||||
Db       8450 CAGATTATTAAACGTTAA 8467


RESULT 4
AF104912
LOCUS       AF104912               16332 bp    DNA     linear   BCT 02-JAN-2001
DEFINITION  Escherichia coli K30 capsule biosynthesis cluster, partial
            sequence.
ACCESSION   AF104912
VERSION     AF104912.2  GI:8469170
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 16332)
  AUTHORS   Drummelsmith,J. and Whitfield,C.
  TITLE     Gene products required for surface expression of the capsular form
            of the group 1 K antigen in Escherichia coli (O9a:K30)
  JOURNAL   Mol. Microbiol. 31 (5), 1321-1332 (1999)
   PUBMED   10200954
REFERENCE   2  (bases 1 to 16332)
  AUTHORS   Drummelsmith,J. and Whitfield,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-1998) Microbiology, University of Guelph, Guelph,
            ON N1G2W1, Canada
REFERENCE   3  (bases 1 to 16332)
  AUTHORS   Drummelsmith,J. and Whitfield,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2000) Microbiology, University of Guelph, Guelph,
            ON N1G2W1, Canada
  REMARK    Sequence update by submitter
COMMENT     On Jun 12, 2000 this sequence version replaced gi:4512003.
FEATURES             Location/Qualifiers
     source          1. .16332
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="E69"
                     /serotype="O9a:K30:H12"
                     /db_xref="taxon:562"
                     /map="45 min; between dcd and gnd"
     misc_feature    338. .396
                     /note="JUMPStart element"
     CDS             672. .2186
                     /note="OrfX; Klebsiella K2 Orf3 homologue"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAD21561.1"
                     /db_xref="GI:4512004"
                     /translation="MPRALSILVFPSFSLRCDSGAENGSCTMIKIARIAVTLGLLSSL
                     GAQAYAAGLVVNDNDLRNDLAWLSDRGVIHLSLSTWPLSQEEIARALKKAKPSYSSEQ
                     VVLARINQRLSALKADFRVTGYTSTDQPGTPQGFGQTQPADNSLGLAFNNSGEWWDVH
                     LQGNVEGGERISNGSRFNANGAYGAVKFWNQWLSFGQVPQWWGPGYEGSLIRGDAMRP
                     MTGFLMQRAEQAAPETWWLRWVGPWQYQISASQMNQYTAVPHAKIIGGRFTFTPFQSL
                     ELGASRIMQWGGEGRPQSFSSFWDGFTGHDNTGTDNEPGNQLAGFDFKFKLEPTLGWP
                     VSFYGQMVGEDESGYLPSANMFLGGVEGHHGWGKDTVNWYVEAHDTRTNMSRTNYSYT
                     HHIYKDGYYQQGYPLGDAMGGDGQLIAGKVELITEDNQRWSTRLVYAKVNPENQSINK
                     AFPHADTLKGLQLGWSGDVYQSVRLNTSLWYTNANNSDSDDVGASAGIEIPFSL"
     gene            2332. .3471
                     /gene="wza"
     CDS             2332. .3471
                     /gene="wza"
                     /note="OMA protein family homologue"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane lipoprotein Wza"
                     /protein_id="AAD21562.1"
                     /db_xref="GI:4512005"
                     /translation="MKKKLVRFSALALAIGFLSGCTIIPGQGLNSLRKNVVELPDSDY
                     DLDKLVNVYPMTPGLIDQLRPEPVIARSNPQLDNLLKSYEYRIGVGDVLMVTVWDHPE
                     LTTPAGQYRSASDTGNWVNSDGTIFYPYIGKVQVAGKTVSQVRQDITSRLTTYIESPQ
                     VDVSIAAFRSQKVYVTGEVANSGKQAITNIPLTVMDAINAAGGLAADADWRNVVLTHN
                     GKDTKISLYALMQKGDLTQNHLLYHGDILFIPSNDDLKVFVMGEVGKQSTLKMDRSGM
                     TLAEALGNAEGISQEMSDATGIFVVRQLKGDRTGKIADIYQLNAQDASAMVLGTEFQL
                     QPYDIVYVTTAPLVRWNRVISQLVPTISGVHDMTETVRYIKRWPN"
     gene            3459. .3905
                     /gene="wzb"
     CDS             3459. .3905
                     /gene="wzb"
                     /codon_start=1
                     /transl_table=11
                     /product="putative acid phosphatase Wzb"
                     /protein_id="AAD21563.1"
                     /db_xref="GI:4512006"
                     /translation="MAKLMFDSILVICTGNICRSPIGERLLRRLLPSKKINSAGVGAL
                     VDHAADESAIRVAEKNGLCLKGHRGTKFTSALARQYDLLLVMEYSHLEQISRIAPEAR
                     GKTMLFGHWLDSKEIPDPYRMSDEAFDSVYQLLEQASKRWAEKLGE"
     gene            3923. .6088
                     /gene="wzc"
     CDS             3923. .6088
                     /gene="wzc"
                     /note="MPA1 protein family homologue"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transmembrane protein Wzc"
                     /protein_id="AAD21564.1"
                     /db_xref="GI:4512007"
                     /translation="MTSVTSKQSTILGSDEIDLGRVIGELIDHRKLIISITSVFTLFA
                     ILYALLATPIYETDALIQIEQKQGNAILSSLSQVLPDGQPQSAPETALLQSRMILGKT
                     IDDLNLQIQIEQKYFPVIARGLAGLMGEKPGNIDITRLYLPDSDDISNNTPSIILTVK
                     DKENYSINSDGIQLNGVVGTLLNEKGISLLVNEIDAKPGDQFVITQLPRLKAISDLLK
                     SFSVADLGKDTGMLTLTLTGDNPKRISHILDSISQNYLAQNIARQAAQDAKSLEFLNQ
                     QLPKVRAELDSAEDKLNAYRKQKDSVDLNMEAKSVLDQIVNVDNQLNELTFREAEVSQ
                     LYTKEHPTYKALMEKRQTLQEEKSKLNKRVSSMPSTQQEVLRLSRDVESGRAVYLQLL
                     NRQQELNIAKSSAIGNVRIIDNAVTDPNPVRPKKTIIIVIGVVLGLIVSVVLVLFQVF
                     LRRGIESPEQLEEIGINVYASIPISEWLTKNARQSGKVRKNQSDTLLAVGNPADLAVE
                     AIRGLRTSLHFAMMEAKNNVLMISGASPSAGKTFISSNLAATIAITGKKVLFIDADLR
                     KGYAHKMFGHKNDKGLSEFLSGQAAAEMIIDKVEGGGFDYIGRGQIPPNPAELLMHPR
                     FEQLLNWASQNYDLIIIDTPPILAVTDAAIIGRYAGTCLLVARFEKNTVKEIDVSMKR
                     FEQSGVVVKGCILNGVVKKASSYYRYGHNHYGYSYYDKK"
```

```
gene            6168. .7598
                /gene="wbaP"
CDS             6168. .7598
                /gene="wbaP"
                /note="glycosyltransferase"
                /codon_start=1
                /transl_table=11
                /product="UDP-Gal::undecaprenolphosphate Gal-1-P
                transferase WbaP"
                /protein_id="AAD21565.1"
                /db_xref="GI:4512008"
                /translation="MTHLARNVSCSITLAVVDFISLTLSLYLALGVLSLTLEHFQQRV
                PQDQIEGWIGLHWLLGVCCVAWYGIRLRHYFYRKTFWFELKEILRTLVIFAIIEVAII
                AFAKWYFSRYVWLMTWVFAAALVPLLRMTAKYILNSLGLWMRDTWIIGNGKNAHEAYK
                AINSENNLGLNVVGFVDVDDNNNGSKNTIDSVQVISVDKTWLTTVDKHTQFIVAVESH
                QSEIRNNWLREFMIRGYRYVSVIPTLRGMPLDSTDMSFIFSHEVMIFRVQQNLAKWSS
                RALKRAFDIVGSLSIIIVLSPVLLYISQKVKKDGGSAIYGHERVGKGGKPFKCLKFRS
                MVVNSKEILERLLENDPKAKAEWDATFKLKDDPRITKIGQFLRRTSLDELPQLFNVLK
                GEMSLVGPRPIITAELERYNDEVDYYLLIKPGMTGLWQVSGRSDVDYDTRVYLDAWYV
                KNWSMWNDIAILFKTIGVVLKKDGAY"
gene            7645. .8877
                /gene="wzy"
CDS             7645. .8877
                /gene="wzy"
                /note="9 putative transmembrane domains"
                /codon_start=1
                /transl_table=11
                /product="repeat unit polymerase Wzy"
                /protein_id="AAD21566.1"
                /db_xref="GI:4512009"
                /translation="MRTLSKAIIILSLIFSVTNYVPISVLVLLLTPISVIGIFYRTDF
                LKFKKIELYLLLLYIYVIGSTLIYNPSSFLEFDFYRKDGNFIISYLILLVFIFLPLNI
                DIDIDKWLKRAFVIFSIASFVAFLIMPKEVPEEGGATVHHFFFLSHNAAGGFYSVIGA
                MALGIYLHSKNKLYLFYAIMFIFFLYMTNSRGSILAIMAAFGYSLIKFKKPFLVFTIF
                LLIQFFIVQETYPTWVSMGKIMSENANYTVSSAAQGIDFQRVGTFIDRLYYLWPRAFD
                NFIHSPLLGLGFGSYDDLYYRYSDVIPHLLAIKEGAMVRHSEAHAHNSTFTILAELGV
                VGYLLFILLFNEINKKIIAIKKVEPGIFMALSLAFWTCIFSSATEHRITTPSQMIPFF
                ILFGICYLKFFGRRNVND"
CDS             8870. .9388
                /note="OrfY"
                /codon_start=1
                /transl_table=11
                /protein_id="AAD21567.1"
                /db_xref="GI:4512010"
                /translation="MIKWINMFYREIRSQPVLKSKVIIFLYRLSSLFFSQSRNPLKYI
                FIIFVFSYYCVVEFLWGIEIKPKTKIGWGIKIFHPNSIIINPGAVLGSGIILRHGVTI
                GNKYNKKTGRETKCPIIKDNVEFGAYACVIGEIVVGEDSIIGAMSYVDFDVEAYSIIS
                SHRGIVLRKKNA"
gene            9381. .10244
                /gene="wcaN"
CDS             9381. .10244
                /gene="wcaN"
                /note="putative glucuronosyltransferase"
                /codon_start=1
                /transl_table=11
                /product="glycosyltransferase WcaN"
                /protein_id="AAD21568.1"
                /db_xref="GI:4512011"
                /translation="MHKTVSIYIPTHNRPKMLARALDSLLTQTYKNFQVLVCNDGSSE
                SYDQIIKEYRSKFSDFIYIENQSPMGACCSRNKLINIADGEYITGLDDDDEFLPSRLT
                DFIDSPNLDKYAYISAGHITKTSSGLFKQKIDERVITLESLLSRNIVGNQVPTKTVFL
                KESGGFDEHLPAWQDYDAWINLTKKIGNGYRIEKYNYQWNIDHEEGRISNSSKAEIGY
                NLFIEKHKDILSKDNLRNLYIQDKINRNVDISLAELFGNFSISILTSVCKYKLKKKFP
                DIKEKIYKVTK"
gene            10241. .11422
                /gene="wcaO"
CDS             10241. .11422
                /gene="wcaO"
                /note="putative galactosyltransferase"
                /codon_start=1
                /transl_table=11
                /product="glycosyltransferase WcaO"
                /protein_id="AAD21569.1"
                /db_xref="GI:4512012"
                /translation="MKVLHISDSYEGGGAEAVLRDTIKVCEELGYQNEIFVSQGKTTI
                ISYVFSIGYYEKLLEKLIDFKPDIVHLHNYYHYLSPSVLYALKKYKKRARIKIIFTAH
                DYHIICPNSGFQYFKDNKRFNFDYNRKESTLFRKFDHRSYAHSLLKVMQHAICYNFLN
                LREVIDIIISPSHFMKSTLSNYGIIKPISVIRNPVSIANTERQILLNDGSIHIVYVGR
                LTPEKGIVEFIKKVNHETTQAIHLHIYGAGESAEEIKSIKCREGFKILFHGFIDRDLL
                ITEISKYHIFVLPSIWLENAPVSIVEAAEAGLPVVVPNYGGLAEMAEETLYNYKFDYE
                DSDLSEVITQAADKKGKNKLNNPDSFSYEMYKESIKKIYAQSFWRMTLKVKSRWENYQ
                S"
CDS             11401. .12651
                /note="OrfZ"
                /codon_start=1
                /transl_table=11
                /protein_id="AAD21570.2"
                /db_xref="GI:8469171"
                /translation="MGKLSVLVLLIVQAFYFSTQVSAVTFSDLNGSNGISSPELSKKQ
                GISIIRIDVPWSEIESVKGNYDYTKYDNEIIRAKKLGYDVLPVLAYAPVWNKSIKGKI
                GSPPLETSSWNGFVESTVKRYRAYPYNITYFQVWNEPTLKAKFWLGNNEDFVNKIYIP
                AAKIIKKNNGKVVFGGWPLSNSISEFDYVLEGLGAIKYTDIIDIHYRGYKSYIKLYDK
                YIKTNKADGIWQSELGYSSKPGNLLSSYLKIMSWALSHNWAKSDTYKIFWYPAWASRE
                KKQYGLATTLKDGTTVLTENGNELVLLNDLYGSGELAVQDVQSNMKSSIDTDDYFFGV
                KISPNRIVIAVLYNGHVDYNQLNYSLMTKNNPRHITLVSSKRKQYQLDHTYEKGVTHI
```

```
Query Match             48.0%;  Score 556;  DB 1;  Length 16332;
Best Local Similarity   68.9%;  Pred. No. 6.6e-100;
Matches  763;  Conservative    0;  Mismatches  345;  Indels    0;  Gaps    0;

Qy         27 TCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTCCGAACAGGTATCATCTGCCATCCT 86
Db      12862 TCTTGTTGCTGATTGGTTAGTAACTTATGCAGGATCGGAAAAAGTTATTGCAGAATTTAT 12921

Qy         87 GCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGTTGATTTTCTAACGGATGAACAAAG 146
Db      12922 AAAACTGTTTCCCCAGTCTGATCTTTATTCAGTTATTGATTTTCTTTCGGATGAATCTCG 12981

Qy        147 AAGACATTTTCTGGGGAAATATGCGACTACCACATTTATTCAAAATTTACCTAAAGCTAA 206
Db      12982 AAGTCGGTTTATGAATAAACGAGCAAAAACAACTTTTATTCAAAAATTACCTAAGGCAAA 13041

Qy        207 AAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCTATTGAACAACTTGATTTATC 266
Db      13042 AAGTAACTACCAGCGATACTTACCATTCATGCCACTTGCTATAGAACAATTGGATGTTTC 13101

Qy        267 AGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCAAAAGGTGTTATTTCCGGACC 326
Db      13102 AAAGCATGAAATCATTCTCTCTAGTAGCCATGCTGTGGCCAAAGGGGTATTAACAGGTCC 13161

Qy        327 AGATCAGCTTCACATTAGCTATGTTCATTCTCCTATTCGATATGCGTGGGATTTACAGCA 386
Db      13162 TGATCAATTGCATATAAGCTATGTTCATTCTCCAATTCGTTATGCTTGGGATTTACAACA 13221

Qy        387 TCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAAGGTTGGTTAGCAAAATGGCT 446
Db      13222 CCAATATTTAAGAGAAGTAGGCCTTGATAAAGGCATTAAATCATATATAGCACGGATGCT 13281

Qy        447 TCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAATGGGGTTGATCATTTTATAGC 506
Db      13282 TTTGCATAAAATTAGATTATGGGATTGCCGTACTGCTAATGGCGTAGATCACTTCATTGC 13341

Qy        507 TAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGGCTTCAGTTAT 566
Db      13342 AAATTCTCAATTCATTGCGCGACGTATCAAAAAAGTATACGGTCGTGATGCTGATGTTAT 13401

Qy        567 ATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAGACTATTATTT 626
Db      13402 TTATCCTCCTGTTGATGTAAATAGATTTGAATTAAATTTTAATAAAGAAGATTATTTTTT 13461

Qy        627 CACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAAGCCTTTAGTAA 686
Db      13462 CACTGCATCTCGATTGGTACCATATAAGAGGATTGATTTGATCGTTGAGGCCTTTTCGGA 13521

Qy        687 AATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAAAAAATAAAGAG 746
```

```
Db      13522 AATGCCCAATCGAAAATTGGTTGTGATTGGGGATGGTCCCGAGATGCAAAAAATTAAAAG 13581

Qy        747 CAAGGCTACAGACAATATAAAATTGCTCGGTTATCAATCTTTTCCTGTTTTAAAAGAGTA 806

Db      13582 TAAGGCTAAGACAAACATTGAAATTTTAGGTTATCAACCTGATTCAGTCATGCAGGAATA 13641

Qy        807 TATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGT 866

Db      13642 TATGCGTAACGCTAAAGCATTTGTTTTTGCTGCAGAAGAGGATTTTGGTATTATACCCGT 13701

Qy        867 CGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAAC 926

Db      13702 AGAGGCTCAGGCTTGCGGGACACCAGTTATTGCATTTGGTAAGGGGGGGGCACTGGAAAC 13761

Qy        927 CGTTCGCCCACTAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGC 986

Db      13762 TATTCGCCCTCTTGGTGTTGATAAACCCACTGGATTATTCTTTCAGAAACAAACTGTGGC 13821

Qy        987 TTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGC 1046

Db      13822 AGATATAATTGAAGCGGTTACTAATTTCGATAATGATTTTGAATCATTTAATCTTCATGA 13881

Qy       1047 TTGTAGAAAAAATGCAGAAAAATTTTCTCGATCAAGATTTGAACAAGAATTTAAGAACTT 1106

Db      13882 TTGCAGGAACAATGCAGAAAAATTCTCTGTGGAAAGATTTAACAACGAAATTTTAAATTA 13941

Qy       1107 TGTTAATGAAAAGTGGAATCTTTTCAAA 1134

Db      13942 TATTACTGACAAGTACGAATCATTCAAA 13969


RESULT 5
AR384270
LOCUS       AR384270               1167 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 999 from patent US 6610836.
ACCESSION   AR384270
VERSION     AR384270.1  GI:40094004
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1167)
  AUTHORS   Breton,G.L. and Osborne,M.
  TITLE     Nucleic acid amino acid sequences relating to Klebsiella pneumoniae
            for diagnostics and therapeutics
  JOURNAL   Patent: US 6610836-A 999 26-AUG-2003;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1167
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             44.9%;  Score 520;  DB 6;  Length 1167;
  Best Local Similarity   65.7%;  Pred. No. 1.7e-92;
  Matches  757;  Conservative    0;  Mismatches  395;  Indels    0;  Gaps    0;

Qy          7 ATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTCCGAA 66

Db         16 AAAACGGACATTAACATTGGTGTCGTGGCTGATTGGTTTATCACTTATGCTGGTTCAGAA 75

Qy         67 CAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGTTGAT 126

Db         76 AAGGTTGTCGCTGAGTTTTTAAGTTTATTTCCTAAATGCGAGCTTTACGCTGTTGTGGAT 135

Qy        127 TTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATTTATT 186

Db        136 TTTCTCTCTCAAGAGAATAAATTACATTTTCAGAATAAAGTAATAACAACTACTTTCATT 195

Qy        187 CAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCT 246

Db        196 CAAAACTTGCCTAAAGCACGTACAAAATACCAGTCATACTTACCATTTATGCCACTGGCA 255

Qy        247 ATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCA 306

Db        256 ATTGAACAGCTTGATGTCTCTAAGCATGATGTAATACTTCTTAGTAGCCATGCAGTCGCT 315

Qy        307 AAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTATTCGA 366

Db        316 AAGGGCGTGCTTACAGGACCTGACCAACTTCACATAAGTTATGTTCATTCTCCTATTCGT 375

Qy        367 TATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAA 426

Db        376 TATGCATGGGATTTGCAACATCAGTACTTGAGAGAGGCTGGATTGAGTAAGGGAGTGAAA 435

Qy        427 GGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAAT 486

Db        436 GCTCTCTTGGCTCGGTGGATATTACATAAAATAAGGATGTGGGATTCCCGGACTGCAAAT 495

Qy        487 GGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATAC 546

Db        496 GGTGTAGACCATTTCATTGCTAATTCTAATTTTATTGCGCGTCGTATAAAAAAAGTCTAT 555

Qy        547 AGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAAT 606

Db        556 GGCAGGAATGCGGATGTTATATATCCTCCTGTTGATGTCGACAACTTCACATTTCAAGAA 615

Qy        607 GAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTT 666

Db        616 GAGAAGGAAGAATATTATTTTACGGCATCTCGTTTGGTTCCGTATAAACGTATTGATTTG 675

Qy        667 ATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCG 726

Db        676 ATTGTTGAGGCATTTAGTTGTATGCCAGGAAGAAAACTTGTTGTTGTCGGTGACGGACCC 735

Qy        727 GAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTATCAATCT 786

Db        736 GAAATGAGAAAGATAAAAAATAAAGCAACAAAAAACATAGATATTCTAGGCTACCAGTCA 795

Qy        787 TTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAG 846

Db        796 AATGAGGCTATGCGGGAGTATATGCAAAAAGCAAAAGCATTTGTTTTTGCGGCCGAAGAA 855

Qy        847 GACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGG 906

Db        856 GATTTTGGCATTACCCCGGTTGAGGCCCAGGCGTGTGGAACACCTGTTATTGCCTTTGGT 915

Qy        907 AAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGCATTTTC 966

Db        916 AAAGGTGGATCTTTGGAAACAGTAAGACCATATGGAGTAAGCAATCCTACTGGTATTTTA 975

Qy        967 TTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAAAATGCA 1026

Db        976 TTTGAAAACCAAGATATACAATCTATAATTAACGGTGTTGATTTATTTGAACGTGTATGC 1035

Qy       1027 TCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCAAGATTT 1086

Db       1036 GATGAAATTAGTCCTGGAGATTGTAGAAAGCATGCTGTAAAATTTTCCATTGATAGATTT 1095

Qy       1087 GAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAACAGATT 1146

Db       1096 CGTAATGAAATAAATGATTATGTTCTAGAAAAATATAACAAATTCTTAGATGCGAAAAAA 1155

Qy       1147 ATTAAACGTTAA 1158

Db       1156 ATAAATTATTAA 1167


RESULT 6
AB041266
LOCUS       AB041266               21411 bp    DNA     linear   BCT 19-SEP-2000
DEFINITION  Actinobacillus actinomycetemcomitans gene cluster for synthesis of
            serotype d-specific polysaccharide antigen, complete cds.
ACCESSION   AB041266
VERSION     AB041266.1  GI:7592804
KEYWORDS    putative mannosyl transferase; putative rhamnosyl transferase;
```

```
            putative O-antigen transporter; Wzz homolog; Unknown.
SOURCE      Actinobacillus actinomycetemcomitans
  ORGANISM  Actinobacillus actinomycetemcomitans
            Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
            Pasteurellaceae; Actinobacillus.
REFERENCE   1  (sites)
  AUTHORS   Nakano,Y., Yoshida,Y., Suzuki,N., Yamashita,Y. and Koga,T.
  TITLE     A gene cluster for the synthesis of serotype d-specific
            polysaccharide antigen in Actinobacillus actinomycetemcomitans
  JOURNAL   Biochim. Biophys. Acta 1493 (1-2), 259-263 (2000)
   PUBMED   10978535
REFERENCE   2  (bases 1 to 21411)
  AUTHORS   Nakano,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2000) Yoshio Nakano, Kyushu University Faculty of
            Dentistry, Department of Preventive Dentistry; 3-1-1 Maidashi,
            Higashi-ku, Fukuoka-shi 812-8582, Japan
            (E-mail:yosh@dent.kyushu-u.ac.jp, Tel:81-92-642-6423,
            Fax:81-92-642-6354)
FEATURES             Location/Qualifiers
     source          1. .21411
                     /organism="Actinobacillus actinomycetemcomitans"
                     /mol_type="genomic DNA"
                     /strain="IDH 781"
                     /db_xref="taxon:714"
     gene            1140. .1934
                     /gene="orf0"
     CDS             1140. .1934
                     /gene="orf0"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA94392.1"
                     /db_xref="GI:7592805"
                     /translation="MSFNINGLRARPHQLEEIINKYQPDVLGLQEIKVSDEVFPYDLV
                     DHLGYHVNHFGQKGHYGVALLSKQAPLSVRKGFPTDDEEAQKRIIMIDLETKFGQLTV
                     INGYFPQGESREHPTKFPAKQKFYADLQRYLEQDHNAQNPVIIMGDMNISPSDLDIGI
                     GEENRKRWLRTGKCSFLPEEREWYQRLYAYGLEDTFRHMNPTVNDKFSWFDYRSKGFD
                     DNRGLRIDHILANKALVAHCVDTGIALDIRAMNKPSDHAPIWAEFK"
     gene            1934. .2722
                     /gene="orf1"
     CDS             1934. .2722
                     /gene="orf1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA94393.1"
                     /db_xref="GI:7592806"
                     /translation="MPLKNPEQNWQNYRTLINGKIASCTVNLDIFETFSPQKYNKVVQ
                     VSLPYEPDENAMPTLDEHHRVITELFKILIQVSALSDVLYAGHIIGDGHLQLHFYCDK
                     PTALFDVLDQFKEHIDHTGVQDDPNWDTYFDFLLPSPLEMKLNVTEEILDMLLQNGRN
                     LADNYLIEHTFQFADEQLMYQFMEETNLSDISFNTMSYSNAPITFPDDEDKESFYIVK
                     IEQEMSLDTDEIFAYVEQFERLAEKFSGEYVGWESDTINQVKAN"
     gene            2832. .3941
                     /gene="wzz"
     CDS             2832. .3941
                     /gene="wzz"
                     /codon_start=1
                     /transl_table=11
                     /product="Wzz homolog"
                     /protein_id="BAA94394.1"
                     /db_xref="GI:7592807"
                     /translation="MTEKVLENDEIDLVELIKVLWNKRLWIILFSLFFTTIAGIYVFT
                     AKEQWTSKAEIIAPRITELGDYFLVKKEYARILGSEFDASMTANNLFGQFNLLSESLD
                     ERKKFFEKSDFYKKESEGKSEREKLQILFDLMTKNIAIKKPDTKKYPDLIGNQISFYA
                     ETPIEAQNVLSKFISTVNISAYQIELDKFLTDFNEIIADLNYEKIKFERDLRIHKSVQ
                     LKNLESALNIAKEAGIKEYSKSFSIADPQLVLQAIAVSEAKVPLSDSKLSDSTYLFML
                     GETYLKAQIDVLKNNNIIYPPRYYQVSSQLSELEPLLIKAKQVKANAFSYQASPDYPV
                     VRDKPKKLLILSIGFILGLILGSIYILFRKLVSNK"
     gene            3941. .5218
                     /gene="orf3"
     CDS             3941. .5218
                     /gene="orf3"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="O-antigen transporter"
                     /protein_id="BAA94395.1"
                     /db_xref="GI:7592808"
                     /translation="MIHIFKNLLESRDNRKLLENFISLFLLQGMNYILPLLLFPYLVR
                     VLGVENFGILSPVTAVIAYFILFTDYGFNITATRDISINKDNNNKVSEIFSSVIIIKL
                     LLGIFSFLVLYVLVYTVQGWKKYNYIYFLTFGSILGQILLPIWFYQGMEKMKFFAIFN
                     VLARIISTILIFLYVKEPEDIYFVPLVNSIAMIISGGTSFLVIRTLFNIRFQIQRFEV
                     LIYYFKNGWNVFLSMFFSNFYRNFNILVLGLLTNDLFVGYYAIAEKLIKIIQVSQDIV
                     GNVLFPYFSEKFIKNNQYFFNFNNKFFPYIVSCYFILWFSTYFLSDQLAQLFNEDVNG
                     SVSLDIKIMSIVILIGGLNYYFGILGLLSMGKNKYFSDCIIITGISNIFISFLFVYFW
                     DDIGASFSLVASEFILLIMIKLKLNKVKLMYKL"
     gene            5252. .6217
                     /gene="orf4"
     CDS             5252. .6217
                     /gene="orf4"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="rhamnosyl transferase"
                     /protein_id="BAA94396.1"
                     /db_xref="GI:7592809"
                     /translation="MIGRIYIIILNYNGWKDTIECLESVLKSDYKNFKIIVVDNNSQN
                     DSIGFICNWAVIPESISAENVYFNKFIRPNAEKPIKYSIWSESEFLQQSNIQEKLILV
                     RANENGGFAKGNNIAIQYAMKDPECRYVWLLNNDTVIKQNTLSKLIEYADKNEIGICG
                     SSLFYYYNPKKVQAYGGHINRFFGTSSHILNVNNIDKKLDYVVGASFLINKRLIEKIG
                     YLPEDYFLYYEETDYCFNARKNGFTLGIALDSIVYHKEGGSTGASQNADKRSEFSDVL
                     TLKNRIKFHKLYLGGGIGLWIGILIAFFNRIRRGQMSRIKSVFID"
     gene            6239. .7534
                     /gene="orf5"
     CDS             6239. .7534
                     /gene="orf5"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA94397.1"
                     /db_xref="GI:7592810"
                     /translation="MIKINMQKVSISNKWKFILVFLIIQLPFQNFFLNQTSLGIIGTN
                     LSIFSIYLSFLFICICDFKHLIFIRLIDIFLLFFLIIISLVGILCSDFTYKDHNLFIK
                     MINNFIIYSQPLIVFLLFNYIFKTKIYLLKTATSTSFYVFIILFLYLILCKFEIFSLD
                     NNKILHGVKNYNMRLRLFSMESSHAITTYITFGLVSFLVCKSLFSKLFILVLLICGSL
                     LIQSKGGILSFILSVLFILFLYITTIKKMMFILFISVITIMLLDITMEDIHVMISSFD
                     EYTSITTRATLIIASLFSLFLYPFGFGFGAYLEFFDKCIIYAIDNVDYIFSMIGLSGN
                     YSEINSYLYDASNYSTKSFFFDLIIYFGWVGLFSFMFYHCEINNKITSKILKVLFVFL
                     VVSQITYVDSVYLYNFWFVFALMYNYWRIYEKKYISNIK"
     gene            7503. .8681
                     /gene="orf6"
     CDS             7503. .8681
                     /gene="orf6"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="mannosyl transferase"
                     /protein_id="BAA94398.1"
                     /db_xref="GI:7592811"
                     /translation="MKKNILVISNDFPYPPNHGGRVDIFERIRAMNNLGHKVYLISTV
                     RSYPNKSEIIFMDNYVYRSFFLSRKTGILPIFSLLPYQVSSRNNIYSLNNIISEVSDI
                     SFDICIIEGHYCLSIFYKLKNRLVINKSYLRVHNDEATYFSQLSKSEKNLLKKVFYKI
                     ESIKFYLYEKKIFSKKEIDALLHISSDELTKYEKKYKINLFLPASINIDDIVLYSIKE
                     KKKVLFVGSLFMPNNIEGLFWYLDEVHSNVLNKFPDYKFIIAGNTKDINVMSLKKKLN
                     EYTSIDFHESPSDLELSKLYYDCSLFINPMLSGAGVKLKTINAIKNGLPVVSTSTGNE
                     GTGLLSSESILISDDPYSFSNNILNLLTNEHERALLTNNAQKFILDKYDLNKNLKKIL
                     "
     gene            8793. .9911
                     /gene="orf7"
     CDS             8793. .9911
                     /gene="orf7"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="mannosyl transferase"
                     /protein_id="BAA94399.1"
                     /db_xref="GI:7592812"
                     /translation="MKKALVHDWFSAFGGAEKCIVSLTNIWNDFDIYSLIDFLNEENR
```

```
                     EIILGGKKAHTSFIQNLPKSKEKYRNYLPLFPFAIEQFNLEDYDLILSSSHCVAKGVL
                     THSNQLHISYVHTPVRYAWDLYFRYLKDNNLDKGLKGILARYFLHKLRIWDLSTINRV
                     DYYIANSHYIARRMKKVYGKEADVIYPPVDVDKFQLCAEKESFYLTASRLVPYKKIDL
                     IVEAFSKTDKKLVVVGNGPDMEKIKSKASKNIEILGYQPDMILLDLMKKAKAFVFAAE
                     EDFGIIPVEAQACGTPVICFNKGGTKETVINNKTGIYFDKQDINSVLFAINEFERKAQ
                     AFEPDSVRENALRFSRQRFEKEIKEYIEDKYQIFRGYK"
     gene            9908. .10933
                     /gene="manB"
     CDS             9908. .10933
                     /gene="manB"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-mannose pyrophosphorylase homolog"
                     /protein_id="BAA94400.1"
                     /db_xref="GI:7592813"
                     /translation="MINLILCGGSGTRLWPLSRVKVPKQFIKLFNHHSLFQLTQSRNA
                     QFSEKTIAVVNDTQFLLAYEQSDNVNQMKFILEPVGRNTAPAIALSCFGLEYNDIVLV
                     TSSDQLIKNQDNYLNMVCKAEKLAEKGFLVTFGLKPKYPETGFGYIETLDEENVLTFR
                     EKPDFDTAKKYVESGRYFWNCGIFCFKAGVFLEELLKYSPDIYYKAKEAYEERVINDN
                     IESISLGNMLEIPSDSIDYAVMEKSDKIKVICGDLGWSDIGSFDSLEKEFAKNQDSNN
                     QIEHHISISSSNNFIWSKKKIVTIDVDDLIIVDTDDVLLVVKKGSSQKVKKVVEYLKE
                     IEPNLVN"
     gene            10941. .12143
                     /gene="orf9"
     CDS             10941. .12143
                     /gene="orf9"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glucosyl transferase"
                     /protein_id="BAA94401.1"
                     /db_xref="GI:7592814"
                     /translation="MKNILLKFILIFVDLIAIFISIALAVTIRKAINLFIDVPEVGYF
                     YTNLRSIYAIQIISFIFLGIYRKRYDFWHETHLIIKGSFLSFIILLAGLALNQNVDIY
                     SRSTLFLAFLFSIFVIPLFKYTAKIFLFKLGFWERKVKIITENENFISEIFSNPYLGY
                     VKSNSSNYDTLFINGTGMGTEKLNKIIEENIRYNRDVIFTPVLSEYDFSKMYICNLFN
                     SRMNIFILDNKLLLISSKIIKFLFDYLITVLSFPFWGLVCLVVAIAIKLEEKKGQLFF
                     KQKRLGIDGKEFTCYKFRTMFEEQSFMAAWLDEHPEEKQYYDIYHKYVNDPRITRIGA
                     FLRKTSLDELPQLFNVLKGDMSLIGPRPYMVNEKDSMGDSMMLVLSMKPGITGLWQVS
                     GRSETKFP"
     gene            13304. .14371
                     /gene="rmlB"
     CDS             13304. .14371
                     /gene="rmlB"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11

  Query Match           34.4%;  Score 398;  DB 1;  Length 21411;
  Best Local Similarity   61.6%;  Pred. No. 9.8e-69;
  Matches  695;  Conservative    0;  Mismatches  415;  Indels   18;  Gaps    3;

Qy         10 AATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTCCGAACAG 69
              ||| |||| |||| | |||| || ||||| ||||| |     | ||  ||| | ||| ||
Db       8787 AATTAAATGAAAAAAGCTCTTGTGCATGACTGGTTTTCTGCGTTTGGTGGCGCGGAAAAG 8846

Qy         70 GTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGTTGATTTT 129
                   |       |     || |||     ||      |  | |||    |  ||||||||
Db       8847 TGTATAGTAAGTTTGACAAATATTTGGAACGATTTTGATATTTATAGTTTAATTGATTTT 8906

Qy        130 CTAACGGATGAACAAAGAAGACATTTTCTGGGG---AAATATGCGACTACCACATTTATT 186
               |||  || ||| | ||   |  | || | ||    ||  | ||   ||| |  |||||
Db       8907 TTAAATGAAGAAAATAGGGAAATTATTTTAGGTGGAAAGAAAGCTCATACGAGTTTTATA 8966

Qy        187 CAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCT 246
              ||||| || |||||  | ||| ||   || |  || |||||||  || | ||  | |||
Db       8967 CAAAACTTGCCTAAGTCAAAAGAAAAATATCGCAATTATTTACCTTTATTTCCTTTTGCT 9026

Qy        247 ATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCA 306
              ||||||||  || || |   ||||  | || | ||  |    |    | ||||  || |||
Db       9027 ATTGAACAGTTTAATCTTGAAGATTATGATTTAATTTTATCTTCATCACATTGTGTCGCA 9086

Qy        307 AAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTATTCGA 366
              || || ||  |  |      || |||| | || ||||| ||||| ||| | ||| ||||
Db       9087 AAGGGCGTATTAACTCATTCAAATCAATTACATATTAGTTATGTGCATACCCCTGTTCGT 9146

Qy        367 TATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAA 426
              ||||| ||||||  |  |   ||  ||| | || ||   ||| ||  |||||||| | |||
Db       9147 TATGCTTGGGATCTTTATTTTCGCTACTTAAAGGATAATAATCTAGATAAAGGACTGAAA 9206

Qy        427 GGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAAT 486
              ||      |||| | ||   || | || ||| |  ||||||||||  |  |||     |||
Db       9207 GGAATACTAGCTAGATATTTTTTACATAAATTGAGAATTTGGGATCTTTCAACGATTAAT 9266

Qy        487 GGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATAC 546
              | || ||| ||| ||| |||||||| || ||||| ||  |  | || |||||||| ||
Db       9267 CGCGTAGATTATTATATTGCTAATTCACATTATATTGCTAGGCGGATGAAAAAAGTTTAT 9326

Qy        547 AGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAAT 606
              ||   |||||     ||||| ||||| || |||||||| ||||| ||| |  |       |
Db       9327 GGAAAAGAGGCAGATGTTATTTATCCTCCGGTAGATGTTGATAAATTTCAGCTTTGTGCT 9386

Qy        607 GAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTT 666
              |||||  ||    | |||||| || || || ||| |||| || || ||     |||||||||
Db       9387 GAAAAAGAATCGTTTTATTTAACGGCTTCTCGTTTGGTTCCTTATAAGAAAATTGATCTT 9446

Qy        667 ATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCG 726
              ||||| || || |||  ||||    | || || || ||||| ||| | ||| |||||||
Db       9447 ATTGTAGAGGCTTTTTCTAAAA---CTGATAAAAAGTTAGTCGTTGTAGGTAATGGACCC 9503

Qy        727 GAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTATCAATCT 786
              || ||| |||| |||||     || || |   | |||||   || |   | || |||||| ||
Db       9504 GATATGGAAAAGATAAAATCTAAAGCAAGTAAGAATATTGAAATTTTAGGATATCAACCT 9563

Qy        787 TTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAG 846
              |   |||||  ||| |  ||| | |   || |  || ||||||||||||| |||||
Db       9564 GATATGATTTTACTAGATTTAATGAAAAAGGCTAAAGCATTTGTTTTTGCAGCTGAAGAA 9623

Qy        847 GACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGG 906
              || ||||| || || ||||| || ||||| ||||| || ||||||||||    ||||
Db       9624 GATTTTGGTATTATTCCTGTTGAGGCTCAGGCTTGTGGGACCCCTGTTATCTGCTTTAAT 9683

Qy        907 AAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGCATTTTC 966
              || || ||  |   ||| ||             ||   |  |  |    || || ||||
Db       9684 AAAGGAGGTACTAAAGAGAC-----------AGTGATTAATAATAAAACAGGTATTTAT 9731

Qy        967 TTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAAAATGCA 1026
              ||  |  ||||  ||||   |  | ||      ||| | | ||||||||||  ||| |||
Db       9732 TTTGATAAACAAGATATAAATAGTGTGTTATTTGCTATAAATGAATTTGAACGAAAAGCA 9791

Qy       1027 TCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCAAGATTT 1086
                  ||||  | |  |    |   | |||||    || ||| ||  |  |||||
Db       9792 CAGGCTTTTGAACCAGATAGTGTGCGTGAGAATGCTTTAAGATTCTCTAGACAACGATTT 9851

Qy       1087 GAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAA 1134
              ||  ||||| | ||| | | | || | || || |    |  |||| | |
Db       9852 GAGAAAGAAATAAAGGAATATATTGAAGATAAATACCAAATTTTTAGA 9899


RESULT 7
AE016853_38
WPCOMMENT
  Sequence split into 64 fragments  LOCUS AE016853 Accession AE016853
     Fragment Name      Begin     End
     AE016853_00            1    110000
     AE016853_01       100001    210000
     AE016853_02       200001    310000
     AE016853_03       300001    410000
     AE016853_04       400001    510000
     AE016853_05       500001    610000
     AE016853_06       600001    710000
     AE016853_07       700001    810000
     AE016853_08       800001    910000
     AE016853_09       900001   1010000
     AE016853_10      1000001   1110000
```

AE016853_11	1100001	1210000
AE016853_12	1200001	1310000
AE016853_13	1300001	1410000
AE016853_14	1400001	1510000
AE016853_15	1500001	1610000
AE016853_16	1600001	1710000
AE016853_17	1700001	1810000
AE016853_18	1800001	1910000
AE016853_19	1900001	2010000
AE016853_20	2000001	2110000
AE016853_21	2100001	2210000
AE016853_22	2200001	2310000
AE016853_23	2300001	2410000
AE016853_24	2400001	2510000
AE016853_25	2500001	2610000
AE016853_26	2600001	2710000
AE016853_27	2700001	2810000
AE016853_28	2800001	2910000
AE016853_29	2900001	3010000
AE016853_30	3000001	3110000
AE016853_31	3100001	3210000
AE016853_32	3200001	3310000
AE016853_33	3300001	3410000
AE016853_34	3400001	3510000
AE016853_35	3500001	3610000
AE016853_36	3600001	3710000
AE016853_37	3700001	3810000
AE016853_38	3800001	3910000
AE016853_39	3900001	4010000
AE016853_40	4000001	4110000
AE016853_41	4100001	4210000
AE016853_42	4200001	4310000
AE016853_43	4300001	4410000
AE016853_44	4400001	4510000
AE016853_45	4500001	4610000
AE016853_46	4600001	4710000
AE016853_47	4700001	4810000
AE016853_48	4800001	4910000
AE016853_49	4900001	5010000
AE016853_50	5000001	5110000
AE016853_51	5100001	5210000
AE016853_52	5200001	5310000
AE016853_53	5300001	5410000
AE016853_54	5400001	5510000
AE016853_55	5500001	5610000
AE016853_56	5600001	5710000
AE016853_57	5700001	5810000
AE016853_58	5800001	5910000
AE016853_59	5900001	6010000
AE016853_60	6000001	6110000
AE016853_61	6100001	6210000
AE016853_62	6200001	6310000
AE016853_63	6300001	6397126

Continuation (39 of 64) of AE016853 from base 3800001 (AE016853 Pseudomonas syringae pv.

Query Match 29.4%; Score 340.6; DB 1; Length 110000;
Best Local Similarity 57.3%; Pred. No. 1.4e-57;
Matches 638; Conservative 0; Mismatches 469; Indels 6; Gaps 1;

```
Qy        3 GGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTC 62
Db    90790 GCAAGGAACTTCCATGAGAATTGCTATCGTCCACGACTGGCTGGTGAGCTACGCTGGCGC 90849

Qy       63 CGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGT 122
Db    90850 CGAGCGCGTGCTTGCTTCGCTGATCAACGTCTGGCCACAGGCTGATCTGTTTTCAGTCAT 90909

Qy      123 TGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATT 182
Db    90910 CGACTTTCTCAGCGATCAGGACCGCGCCCACCTGCGCGGCAAGGTAGCCCGAACAACGTT 90969

Qy      183 TATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACT 242
```

```
Db    90970 CATCCAGAAACTGCCCGGTGCCCGCAAACACTATTCTCGCTACCTGCCGCTGATGCCGCT 91029

Qy      243 GGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGT 302
Db    91030 GGCCATCGAACAACTGGACCTGTCCGGCTATGACCTGATCATCAGCAGCAGCCATGCCGT 91089

Qy      303 TGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTAT 362
Db    91090 GGCCAAAGGCGTGCTCAGCGGCCCGGACCAACTGCATATCAGCTATGTGAATTCGCCCAT 91149

Qy      363 TCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAAT 422
Db    91150 CCGCTATGCCTGGGACTTGCAGCACCAGTACCTGCGCGAGTCCGGGCTGGACAAGGGAAT 91209

Qy      423 TAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGC 482
Db    91210 CAAGAGCAGCCTGGCGCGCCTGGTCCTGCACTACATGCGCCTCTGGGATCAGCGCACTGC 91269

Qy      483 AAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGT 542
Db    91270 GGCGGGGGTAGATGCATTCATCGCCAATTCGGCGTTCATCGGTGCGCGCATCAACAAGGC 91329

Qy      543 ATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAA 602
Db    91330 GTATCGCCGGGATTCGACGGTCATCTACCCGCCCGTGGACACACTGGCCTTCACCCCGCT 91389

Qy      603 AAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGA 662
Db    91390 GGGTGCCCGGCAGGACTTCTATCTCAGCGCCTCGCGCATGGTGCCTTACAAACGCATGCC 91449

Qy      663 TCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGG 722
Db    91450 GATGATCGTCGAAGCCTTCGCTGCCATGCCCGACAAACGTCTGATCATGATCGGTGACGG 91509

Qy      723 ACCGGAGATGAAAAAAATAAAGAGCAAGGC------TACAGACAATATAAAATTGCTCGG 776
Db    91510 GCCGGACCTCGCCAAGGCGCAGGCCATTGCGCGTCTGGCACCCAATGTCACGCTGCTGGG 91569

Qy      777 TTATCAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGC 836
Db    91570 TTTCCAGCCAGCCGCTGTCCTGCTGGAACACATGCGCAGCGCCAAAGCCTTTGTCTTTGC 91629

Qy      837 AGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTAT 896
Db    91630 TGCTGAGGAAGACTTCGGCATCAGCCCGGTTGAAGCGCAGGCCTGCGGTACGCCGGTGAT 91689

Qy      897 TGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGAC 956
Db    91690 CGCGTTTGGCAAGGGCGGCGTGCTGGAGACGGTGCTCGGCCTGGATCACATGCACCCGAC 91749

Qy      957 TGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGA 1016
Db    91750 CGGCGTGCTCTACCGACAGCAGAACGTAGAGTCGCTGACCGCCGCCATCGGTGAATTCGA 91809

Qy     1017 AAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCG 1076
Db    91810 AGCCGCGCAATCACGCATCACACCACAGGCCTGTCGATCCAATGCCGAGCGCTTCTCGGT 91869

Qy     1077 ATCAAGATTTGAACAAGAATTTAAGAACTTTGT 1109
Db    91870 CGCGCGCTTCGAGCAGGAGATAAAGGCGTTCGT 91902
```

RESULT 8
CP000075_38
WPCOMMENT
Sequence split into 61 fragments LOCUS CP000075 Accession CP000075

Fragment Name	Begin	End
CP000075_00	1	110000
CP000075_01	100001	210000
CP000075_02	200001	310000
CP000075_03	300001	410000
CP000075_04	400001	510000

CP000075_05	500001	610000
CP000075_06	600001	710000
CP000075_07	700001	810000
CP000075_08	800001	910000
CP000075_09	900001	1010000
CP000075_10	1000001	1110000
CP000075_11	1100001	1210000
CP000075_12	1200001	1310000
CP000075_13	1300001	1410000
CP000075_14	1400001	1510000
CP000075_15	1500001	1610000
CP000075_16	1600001	1710000
CP000075_17	1700001	1810000
CP000075_18	1800001	1910000
CP000075_19	1900001	2010000
CP000075_20	2000001	2110000
CP000075_21	2100001	2210000
CP000075_22	2200001	2310000
CP000075_23	2300001	2410000
CP000075_24	2400001	2510000
CP000075_25	2500001	2610000
CP000075_26	2600001	2710000
CP000075_27	2700001	2810000
CP000075_28	2800001	2910000
CP000075_29	2900001	3010000
CP000075_30	3000001	3110000
CP000075_31	3100001	3210000
CP000075_32	3200001	3310000
CP000075_33	3300001	3410000
CP000075_34	3400001	3510000
CP000075_35	3500001	3610000
CP000075_36	3600001	3710000
CP000075_37	3700001	3810000
CP000075_38	3800001	3910000
CP000075_39	3900001	4010000
CP000075_40	4000001	4110000
CP000075_41	4100001	4210000
CP000075_42	4200001	4310000
CP000075_43	4300001	4410000
CP000075_44	4400001	4510000
CP000075_45	4500001	4610000
CP000075_46	4600001	4710000
CP000075_47	4700001	4810000
CP000075_48	4800001	4910000
CP000075_49	4900001	5010000
CP000075_50	5000001	5110000
CP000075_51	5100001	5210000
CP000075_52	5200001	5310000
CP000075_53	5300001	5410000
CP000075_54	5400001	5510000
CP000075_55	5500001	5610000
CP000075_56	5600001	5710000
CP000075_57	5700001	5810000
CP000075_58	5800001	5910000
CP000075_59	5900001	6010000
CP000075_60	6000001	6093698

Continuation (39 of 61) of CP000075 from base 3800001 (CP000075 Pseudomonas syringae pv.

Query Match 29.4%; Score 340.2; DB 1; Length 110000;
Best Local Similarity 57.2%; Pred. No. 1.7e-57;
Matches 640; Conservative 0; Mismatches 473; Indels 6; Gaps 1;

```
Qy          3 GGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTC 62
              | ||  ||     || | |||  || |||| || || ||| |  | ||||| || ||| |
Db      68818 GCAAGGAACGCCCATGAGAATCGCTATCGTCCACGACTGGCTGGTGAGTTACGCCGGCGC 68877

Qy         63 CGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGT 122
              ||| |  ||      |  |  |  | || |  || |  ||| |  | | | ||||  |
Db      68878 CGAGCGCGTGCTCGCGTCACTGATCAACGTCTGGCCCGCTGCGGACCTGTTTGCGGTCAT 68937

Qy        123 TGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATT 182
              || || || |  ||| |  |  |    || | |  || ||    || |  || || ||
Db      68938 CGACTTCCTCAGCGATCAGGACCGCGCGCACCTGCATGGCAAGGTCGCCAGAACGACGTT 68997

Qy        183 TATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACT 242
              || || ||  | ||     ||     |||   ||   |   ||  | || || ||||| ||
Db      68998 CATCCAGAAGCTGCCCGGCGCGCGCAAACACTATTCGCGCTACCTGCCGCTGATGCCGCT 69057

Qy        243 GGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGT 302
              ||| || || ||||| ||| | || |    | |  | ||||| || |||  |||| | ||
Db      69058 GGCCATCGAGCAACTGGATCTGTCCGGCTATGACCTGATCATCAGCAGCAGCCATGCGGT 69117

Qy        303 TGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTAT 362
              || || || ||  | | ||| || || || || ||||| |||||||| || || || ||
Db      69118 GGCCAAGGGCGTGCTGTGCGGCCCGGACCAACTGCACATCAGCTATGTGCACTCGCCGAT 69177

Qy        363 TCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAAT 422
              || ||||| ||||| || ||||| |||||||| | || ||    |||||| || |||
Db      69178 CCGCTATGCCTGGGACTTGCAGCACCAGTACCTCCAGGAATCAGGGCTGAACAAGGGAGC 69237

Qy        423 TAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGC 482
              || ||  |  | |||     | | || ||  | || ||  | ||||||    || ||  |
Db      69238 CAAGGGCGGTCTGGCACGCCTGATCCTGCATTACATGCGCCTGTGGGATCAGCGCACATC 69297

Qy        483 AAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGT 542
              |  || || ||    || || || || ||    | |||||       | || |  || |
Db      69298 GACCGGCGTGGACGCCTTCATTGCCAACTCCGGGTTTATCGGCGCGCGCATCACCAAGGC 69357

Qy        543 ATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAA 602
              ||  | || ||  |  | || || || |||| ||||    |       ||    |  |
Db      69358 CTATCGTCGCGACTCCACCGTCATCTACCCGCCGGTAGACACGCTCGGCTTCACCGCCCA 69417

Qy        603 AAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGA 662
                |      |||| ||| ||   || || || |||| || |||||||| ||
Db      69418 GGGCGCGCGCGGGGACTTCTATCTCTGTGCCTCGCGAATGGTGCCTTACAAACGCATGCC 69477

Qy        663 TCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGG 722
              | || ||||||||||||      ||||| || ||    |  |  | || || || ||
Db      69478 GATGATCGTCGAAGCCTTTGCCGCCATGCCTGACAAACGCCTGATCATGATCGGCGACGG 69537

Qy        723 ACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGA------CAATATAAAATTGCTCGG 776
              |||||  ||   ||      ||  ||  ||.|   |      |||  | |   |||| ||
Db      69538 GCCGGACCTGGCCAAGGCGCAGGCCATTGCCAGCCAGGTGAGCAACGTGACGCTGCTGGG 69597

Qy        777 TTATCAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGC 836
              || |||  |      ||  |   ||| |||||  |||||||||||||||| || ||
Db      69598 TTTTCAGCCAGGCAGCGTGCTGCTCGAGCATATGCGCAGCGCCAGGGCGTTTGTCTTCGC 69657

Qy        837 AGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTAT 896
              || || ||||| || || || |  || ||||| || || ||||||||||| || || ||
Db      69658 CGCCGAGGAGGATTTCGGCATCAGCCCGGTCGAGGCCCAGGCTTGCGGTACGCCGGTCAT 69717

Qy        897 TGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGAC 956
              ||||| |  ||||| || | | || ||||| |||   || | |      | | |||||
Db      69718 CGCCTTCGCCAAGGGCGGCGTCATGGAGACCGTGCGCGGCCTGGATCACCCGCAGCCGAC 69777

Qy        957 TGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGA 1016
              ||| | |||| |  || || |  | || ||  ||   | ||  |  | ||||| ||
Db      69778 CGGCGTGTTCTACCGAGAGCAAACGGTAGCGTCACTGATCGCCGCCATCGGCGAATTCGA 69837

Qy       1017 AAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCG 1076
              |      ||||    |  |  |  |||| || |    ||||| ||    |||||
Db      69838 AGCGGCGCAATCACGCATCTCTCCCGAGGCCTGCCGGGCCAATGCCGAGCGGTTTTCGGT 69897

Qy       1077 ATCAAGATTTGAACAAGAATTTAAGAACTTTGTTAATGA 1115
                |  | || || || ||| | ||   ||| ||  | ||
Db      69898 CGCGCGTTTCGAGCAGGAAATCAAAGCCTTCGTCGAAGA 69936
```

RESULT 9
CP000058_36
WPCOMMENT
Sequence split into 60 fragments LOCUS CP000058 Accession CP000058

Fragment Name	Begin	End
CP000058_00	1	110000
CP000058_01	100001	210000

CP000058_02	200001	310000
CP000058_03	300001	410000
CP000058_04	400001	510000
CP000058_05	500001	610000
CP000058_06	600001	710000
CP000058_07	700001	810000
CP000058_08	800001	910000
CP000058_09	900001	1010000
CP000058_10	1000001	1110000
CP000058_11	1100001	1210000
CP000058_12	1200001	1310000
CP000058_13	1300001	1410000
CP000058_14	1400001	1510000
CP000058_15	1500001	1610000
CP000058_16	1600001	1710000
CP000058_17	1700001	1810000
CP000058_18	1800001	1910000
CP000058_19	1900001	2010000
CP000058_20	2000001	2110000
CP000058_21	2100001	2210000
CP000058_22	2200001	2310000
CP000058_23	2300001	2410000
CP000058_24	2400001	2510000
CP000058_25	2500001	2610000
CP000058_26	2600001	2710000
CP000058_27	2700001	2810000
CP000058_28	2800001	2910000
CP000058_29	2900001	3010000
CP000058_30	3000001	3110000
CP000058_31	3100001	3210000
CP000058_32	3200001	3310000
CP000058_33	3300001	3410000
CP000058_34	3400001	3510000
CP000058_35	3500001	3610000
CP000058_36	3600001	3710000
CP000058_37	3700001	3810000
CP000058_38	3800001	3910000
CP000058_39	3900001	4010000
CP000058_40	4000001	4110000
CP000058_41	4100001	4210000
CP000058_42	4200001	4310000
CP000058_43	4300001	4410000
CP000058_44	4400001	4510000
CP000058_45	4500001	4610000
CP000058_46	4600001	4710000
CP000058_47	4700001	4810000
CP000058_48	4800001	4910000
CP000058_49	4900001	5010000
CP000058_50	5000001	5110000
CP000058_51	5100001	5210000
CP000058_52	5200001	5310000
CP000058_53	5300001	5410000
CP000058_54	5400001	5510000
CP000058_55	5500001	5610000
CP000058_56	5600001	5710000
CP000058_57	5700001	5810000
CP000058_58	5800001	5910000
CP000058_59	5900001	5928787

Continuation (37 of 60) of CP000058 from base 3600001 (CP000058 Pseudomonas syringae pv.

Query Match 28.7%; Score 332.6; DB 1; Length 110000;
Best Local Similarity 56.9%; Pred. No. 5.2e-56;
Matches 633; Conservative 0; Mismatches 474; Indels 6; Gaps 1;

```
Qy          3 GGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTC 62
              | ||  ||     || | ||| || | || || || ||| |  | |||||| || ||  |
Db      52773 GCAAGGAACGCCCATGAGAATTGCTATTGTCCACGACTGGCTGGTGAGTTACGCCGGTGC 52832

Qy         63 CGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGT 122
              ||| |  ||     || || |  |  | | | || || ||| |  | | | | ||  |
Db      52833 CGAGCGCGTGCTTGCTTCCCTGATAAACATCTGGCCCGATGCGGATCTGTTTGCAGTCAT 52892

Qy        123 TGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATT 182
```

```
                || ||||| |  ||| |  |  |    ||  | |  || ||    || |  || || ||
Db      52893 CGACTTTCTCAGCGATCAGGACCGCGCGCAGCTGCGCGGCAAGGTAGCCAGAACAACGTT 52952

Qy        183 TATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACT 242
              || || ||  | ||     ||    |||  |||         ||  | ||| || ||||| ||
Db      52953 CATCCAGAAGCTGCCCGGCGCGCGCAAACATTATTCACGCTACCTGCCGCTGATGCCGCT 53012

Qy        243 GGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGT 302
              ||| || || ||||| ||  | || |       |  | ||||| || |||  |||  | ||
Db      53013 GGCCATAGAGCAACTGGACCTCTCCGGCTACGACCTGATCATCAGCAGCAGCCACGCAGT 53072

Qy        303 TGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTAT 362
              || || || ||  ||| ||| || || || || ||||| |||||||| ||||| || ||
Db      53073 GGCCAAGGGCGTGCTTTGCGGCCCGGACCAACTGCACATCAGCTATGTGCATTCGCCAAT 53132

Qy        363 TCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAAT 422
              ||| || || ||||| || ||||| ||||||  | | || ||    |||    || |||||
Db      53133 TCGCTACGCCTGGGACTTGCAGCACCAGTACTTGCAGGAATCAGGGCTGGGCAAGGGAAT 53192

Qy        423 TAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGC 482
              || ||  |  | ||       | | || ||  | || ||  | |||||     || ||  |
Db      53193 CAAGGGCAGTCTGGCGCGCCTGGTCCTGCATTACATGCGCCTGTGGGACCAACGTACGTC 53252

Qy        483 AAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGT 542
              |  || || ||    || || || |||||    | |||||    | ||    || |
Db      53253 GACCGGCGTGGACGCATTCATTGCCAATTCGGGCTTTATCGGCGCACGCATCTCCAAGGC 53312

Qy        543 ATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAA 602
              ||  |||||||  |  | || || |||||||| || ||    |     |||         |
Db      53313 GTATCGACGTGACTCCACCGTCATCTATCCGCCAGTGGACACGCTCGGTTTCACCCCCCA 53372

Qy        603 AAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGA 662
              |   |          ||||  ||  |    ||||| || ||||| |||||||| || ||
Db      53373 AGGCGCGCGTGGGGACTTCTACCTGTGTGCATCGCGGATGGTGCCCTACAAGCGCATGCC 53432

Qy        663 TCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGG 722
              | || ||||||||||||          |||||||| ||       |  |  | || ||||| ||
Db      53433 AATGATCGTCGAAGCCTTCGCCGCCATGCCGGACAAACGCCTGATCATGATCGGTGACGG 53492

Qy        723 ACCGGAGATGAAAAAAATAAAGAGCAAGGCTA------CAGACAATATAAAATTGCTCGG 776
              ||||||  ||  ||     ||  ||  || |      |   |||  |||    |||| ||
Db      53493 ACCGGACCTGGCCAAGGCGCAGGCCATTGCCAGGCAGGCGCCCAACGTAACGCTGCTGGG 53552

Qy        777 TTATCAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGC 836
              ||  ||  ||     | ||  |    ||  |||||| ||||||| ||||||| || || ||
Db      53553 TTTCCAGCCTGGCGCGGTACTGCTGGAACATATGCGCAGCGCCAAGGCGTTCGTCTTCGC 53612

Qy        837 AGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTAT 896
              || || ||||| || || || |  || || ||||| ||||| ||||| || || || ||
Db      53613 CGCCGAGGAGGATTTCGGCATCAGCCCGGTTGAAGCCCAAGCCTGCGGCACGCCGGTAAT 53672

Qy        897 TGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGAC 956
              || || || ||||| ||   | | || ||||| | |   || | |  | | |||||||
Db      53673 CGCGTTCGGCAAGGGCGGCATCATGGAGACCGTGCACGGCCTGGATCAACCGCAACCGAC 53732

Qy        957 TGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGA 1016
              ||| |  ||| |  | | || |   | ||  |  ||   |  ||  |  | || |||||
Db      53733 CGGCGTGCTCTACCGGCAGCAAACCGTGGCAGCGCTGAGCGCCGCCATCGGCGAGTTTGA 53792

Qy       1017 AAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCG 1076
              |     ||||  |  | | ||||| ||  |    || || |||  |||||
Db      53793 AGCTGCGCAATCACGCATCTCGCCGCAGGCCTGCCGGGCCAACGCCGAACGGTTTTCGGT 53852

Qy       1077 ATCAAGATTTGAACAAGAATTTAAGAACTTTGT 1109
                |  | || || || ||  | ||  ||| ||
Db      53853 GGCGCGTTTCGAGCAGGAGATCAAAGCCTTCGT 53885
```

RESULT 10
AE009256
LOCUS AE009256 8375 bp DNA linear BCT 28-MAY-2004
DEFINITION Agrobacterium tumefaciens str. C58 linear chromosome, section 26 of 187 of the complete sequence.

```
ACCESSION   AE009256 AE008689
VERSION     AE009256.1  GI:17741647
KEYWORDS    .
SOURCE      Agrobacterium tumefaciens str. C58
  ORGANISM  Agrobacterium tumefaciens str. C58
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Agrobacterium.
REFERENCE   1  (bases 1 to 8375)
  AUTHORS   Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida,N.F. Jr.,
            Zhou,Y., Bovee,D. Sr., Chapman,P., Clendenning,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenthner,D., Kutyavin,T., Levy,R., Li,M.,
            McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
            and Nester,E.W.
  TITLE     The genome of the natural genetic engineer Agrobacterium
            tumefaciens C58
  JOURNAL   Science 294 (5550), 2317-2323 (2001)
   PUBMED   11743193
REFERENCE   2  (bases 1 to 8375)
  AUTHORS   Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida,N.F. Jr.,
            Zhou,Y., Bovee,D. Sr., Chapman,P., Clendenning,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenthner,D., Kutyavin,T., Levy,R., Li,M.,
            McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
            and Nester,E.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2001) Department of Microbiology, University of
            Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
            98195-7242, USA
FEATURES             Location/Qualifiers
     source          1. .8375
                     /organism="Agrobacterium tumefaciens str. C58"
                     /mol_type="genomic DNA"
                     /strain="C58"
                     /isolate="U. Washington"
                     /db_xref="taxon:176299"
     gene            complement(79. .1302)
                     /gene="Atu3266"
     CDS             complement(79. .1302)
                     /gene="Atu3266"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAL44082.1"
                     /db_xref="GI:17741648"
                     /translation="MTSGEQAKTPLQAPILLTNVKPVGFGKGASQSSTDILIGGDGKI
                     AAVGSALQAPADTQRIDAKGAFISPGWVDLHVHIWHGGTDISIRPSECGAERGVTTLV
                     DAGSAGEANFHGFREYIIEPSRERIKAFLNLGSIGLVACNRVPELRDIKDIDLDRILE
                     CYAENSEHIVGLKVRASHVITGSWGVTPVKLGKKIAKILKVPMMVHVGEPPALYDEVL
                     EILGPGDVVTHCFNGKSGSSIMEDEDLFNLAERCAGEGIRLDIGHGGASFSFKVAEAA
                     IARGLLPFSISTDLHGHSMNFPVWDLATTMSKLLSVDMPFENVVEAVTRNPASVIRLD
                     MENRLDVGQRADFTVFDLVDADLEATDSNGDVSRLKRLFEPRYAVIGAEAIAASRYIP
                     RARKLVRHSHGYSWR"
     gene            complement(1304. .2179)
                     /gene="Atu3267"
     CDS             complement(1304. .2179)
                     /gene="Atu3267"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter, membrane spanning protein
                     [dipeptide]"
                     /protein_id="AAL44083.1"
                     /db_xref="GI:17741649"
                     /translation="MADITSTAPPTRNPTILFIRRLLKRRTVAAGLIVLLVFVLLAVF
                     APMIAPYSPSKLSIVNRLKPPSELYWFGTDEFGRDVFSRTIYAGRLSLLVGAAVVALS
                     SIIGITLGLLAGFFQKLDTPIARLIDAMMAFPDILLAIALVAALGPSLTTVIIALAVV
                     YSPRLARIVRASTLVIRELPYVEAARALGISTFHIMTRHVLRNLVSPILVQCTFLFAS
                     AMLAEAGLSFLGLGVSPEIPTWGTMISAGRQYIGQADWMTYFPGFAIILSVLSLQMVG
                     DGLRDMLDPRLRRDL"
     gene            complement(2179. .3087)
                     /gene="Atu3268"
     CDS             complement(2179. .3087)
                     /gene="Atu3268"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter, membrane spanning protein
                     [dipeptide]"
                     /protein_id="AAL44084.1"
                     /db_xref="GI:17741650"
                     /translation="MIVVMFLVVTIVFVILRVTPGDPAAVMLGPDATAQDIAELRTRL
                     GLDQSLGVQYVYYIGQLLKGDLGQSIFLNMPVGSALLDRAEPTFFLTIFSLLIAGVIA
                     LPIGIYAAYRRGSFVDQAATTIAMLAASIPSFWLGLILMQFFAVRLDLFPVSGYGGPG
                     ASFLERMYHLTLPAIALGLVSSALILRFTRASMLDVLGDDYVRTARAKGVKERRVVMK
                     HALKNALIPILTVLGLTAAVLISGAVVTETVFGLPGVGNLVVSAVLRRDYPVIQGALL
                     VIAGLYVLINFAIDMLYLLVDPRVRY"
     gene            complement(3450. .4976)
                     /gene="Atu3269"
     CDS             complement(3450. .4976)
                     /gene="Atu3269"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter, substrate binding protein
                     [dipeptide]"
                     /protein_id="AAL44085.1"
                     /db_xref="GI:17741651"
                     /translation="MRRKIAFLATAALLLAPGVMTAAEKGGVINVATIGEPPTLDPMA
                     STADLVGIVTQHIFETLYTFDKGWKVTPLLAESLPEISADGKTYTIKLRQGIKFHDGS
                     DMTSEDVVASLGRWMKIASRGKQAAGFIDNISAPDAGTVTITLKQPYAPLTSLLAFNN
                     SAAIILPAEKQAEPMAEFVGTGPYMLKERKADQYIQLVRFDGYKSRDGESDGYGGARH
                     QYLDEIRFVPVPDPNTRVEAAVSGQYDYVDSIPVESFEKVKSSSASQPMMLKPFGFPV
                     FVFNTKEGLASKLEVRKAVTEALNMEDMLAAAFGSTDFYALDGAYYPDNFSWHTEEGV
                     DGNYNVGNPEAAAEKLKAAGYNGEPLRILTSRQYEFHYKMAQVAAEYLKLAGFKVDMQ
                     VVDWATLTQRRADPKLWDIYISHSPFLPEPALMGALSTSSPGWWDTPARKAAVDAFTS
                     EADPEKRLALWANVQKTIYAELPLIKIGDFNAAAAKSNKLEGVDPAPWPYFWNASVTK
                     "
     gene            complement(5036. .5278)
                     /gene="Atu3270"
     CDS             complement(5036. .5278)
                     /gene="Atu3270"
                     /note="putative; ORF located using Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAL44086.1"
                     /db_xref="GI:17741652"
                     /translation="MGGYVCPSPIFNNTSFANAKQPSCVILRRRFPRECVQQYPPCRL
                     APFCSPEKSCFQFADTGNPSSRRKVIDKASYRDILL"
     gene            5415. .6539
                     /gene="Atu3271"
     CDS             5415. .6539
                     /gene="Atu3271"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyltransferase"
                     /protein_id="AAL44087.1"
                     /db_xref="GI:17741653"
                     /translation="MKTAIVHDWLTDRGGAEKVLHNLLEIYPHADLYCLVDFMSERDR
                     GFLGGRPVNTSFIQKLPFARKKYRSYLPFMPLAIEQFDLSGYDLVISSSYAVAKGVIT
```

```
                     GPGQFHVSYIHSPIRYAWDLQHQYLKETRLNRGFASWIARAILHYIRMWDIRTTHGVD
                     LMTVNSKFIRSRVRKCYGRDSTVIYPPVDTAHLSPSDRKGDYFVTASRLVPYKKVHLI
                     VEAFAKMPDKRLVVIGDGPDMKRIRDIATPNVEILGFVGNEELHKYMAEAKAFVFAAE
                     DDPGIVPVESQALGTPVIAYGKGGVLETVIPLGASNQPTGLYFGEQTPEAICAAVQEF
                     VDNSDRFDKSACVDNAARFSRERFLREFADTVNFALAERA"
     gene            6947. .8212
                     /gene="pssN"
                     /note="synonym: Atu3272"
     CDS             6947. .8212
                     /gene="pssN"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer"
                     /codon_start=1
                     /transl_table=11
                     /product="exopolysaccharide export protein"
                     /protein_id="AAL44088.1"
                     /db_xref="GI:17741654"
                     /translation="MFNNLYGVRTIFTFSDGIYFEDQMRLLNVSILLASAALSGCTVT
                     AASGPDAATIESNASVKFASKDKHKTAGVDYALIDVNSAVLNYVGDTTNTTLLGSFGG
                     GKGGVPALPMGVGDVVQVAIFESQAGGLFIPSEAGSRPGNFISLPNQTVDREGNISVP
                     YAGKIRATGRNVENVQTEIEERLANRAIEPQVLITKISSRSAQASVLGDVKEPTKVQL
                     SEAGDRVLDVISYAKGLSAPNIESYVTLIRRGKTARINYNHLVSTPSENIYVVPGDTI
                     MVERERRTYLAFGASGLNGRFEFEDASLKLSDALGKAGGLLDSRADPAQVYIYRTVKR
                     DLLVKLGIDTSKFPGQEIPVIFRANLRDPSTFFATTKFPMQDRDIIYVTNSQATELYK
                     FLDLVGSVPATTGNVAEDVLTTRNAIRAF"
ORIGIN

  Query Match             23.2%;  Score 269;  DB 1;  Length 8375;
  Best Local Similarity   53.6%;  Pred. No. 3.7e-43;
  Matches  582;  Conservative    0;  Mismatches  500;  Indels    3;  Gaps    1;

Qy         16 ATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTCCGAACAGGTATCA 75
              || || |   |  |||| ||||| ||| |       |  ||  ||||  | ||
Db       5415 ATGAAGACGGCAATCGTGCATGACTGGCTCACCGACCGCGGTGGGGCCGAGAAAGTGCTT 5474

Qy         76 TCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGTTGATTTTCTAACG 135
                | | ||  |  ||| |||| | ||| |  | ||||   | || |||||  | |
Db       5475 CACAACCTTCTCGAAATTTATCCTCACGCGGACCTTTATTGCCTCGTCGATTTCATGAGC 5534

Qy        136 GATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATTTATTCAAAATTTA 195
              ||     |  |  |  | |   |||       |  | ||| | || || || ||  |
Db       5535 GAGCGCGACCGGGGCTTTCTCGGCGGGCGTCCCGTCAACACCTCCTTCATCCAGAAGCTG 5594

Qy        196 CCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCTATTGAACAA 255
              ||   ||    ||   || |     ||| | ||  | ||||| || || || || ||
Db       5595 CCCTTCGCCCGGAAGAAATATCGCTCCTATCTGCCGTTTATGCCGCTCGCCATCGAGCAG 5654

Qy        256 CTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCAAAAGGTGTT 315
              | ||| |    | |  | |  || ||||      |  | || | || || || || ||
Db       5655 TTCGATCTTAGCGGTTATGACCTCGTCATCTCCTCCAGCTATGCGGTCGCCAAGGGCGTC 5714

Qy        316 ATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTATTCGATATGCGTGG 375
              ||  |||| || |  ||| ||||| |    |||  | ||     || ||||| || || |||
Db       5715 ATCACCGGTCCCGGCCAGTTTCACGTCTCCTACATCCACAGCCCCATTCGCTACGCCTGG 5774

Qy        376 GATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAAGGTTGGTTA 435
              ||| | |||||||| || || || ||  |    || |||   ||  |       ||| ||
Db       5775 GATCTCCAGCATCAATATCTCAAGGAAACCCGGCTCAATCGCGGCTTCGCCTCGTGGATT 5834

Qy        436 GCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAATGGGGTTGAT 495
              ||       |||||||| | || || || |||||   || ||| |  |||| || |||
Db       5835 GCGCGCGCCATTCTTCACTACATTCGCATGTGGGATATCCGCACCACCCATGGCGTCGAT 5894

Qy        496 CATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAG 555
              |   | |  |  |||||  |||  |||     |  |  |    |||   ||  | || ||
Db       5895 CTGATGACGGTGAATTCCAAATTCATCCGCAGCCGCGTGCGCAAATGTTATGGGCGCGAT 5954

Qy        556 GCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAA 615
              | | || || || ||||| || |||    |   || ||           |    ||
Db       5955 TCCACGGTGATCTACCCGCCGGTCGATACCGCGCATCTTTCCCCTTCCGACAGGAAAGGC 6014

Qy        616 GACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAA 675
              || ||||   |||| || || ||  |||| ||||||||     |  | || || || ||
```

```
Db       6015 GATTATTTCGTCACCGCCTCACGGCTGGTGCCCTACAAGAAAGTGCACCTGATCGTGGAG 6074

Qy        676 GCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAA 735
              ||||||   || ||||| || ||     | || || || || || || ||||| ||||||
Db       6075 GCCTTTGCGAAGATGCCCGATAAACGGCTTGTCGTGATCGGCGACGGGCCGGACATGAAA 6134

Qy        736 AAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTATCAATCTTTTCCTGTT 795
                 ||       ||  || ||    ||| |  || | |||||||| |         |  |
Db       6135 CGCATCCGCGACATCGCCACGCCTAATGTCGAAATTCTCGGTTTTGTCGGCAATGAGGAG 6194

Qy        796 TTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGA 855
              |  |  | || |||  |   |||| ||| || |||||||| || ||||| || || ||
Db       6195 CTTCACAAATACATGGCGGAAGCCAAGGCATTCGTTTTTGCCGCTGAAGATGATTTCGGC 6254

Qy        856 ATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGG 915
              ||  | || |||||  | || ||    || ||||| || || |||| ||| ||||| ||
Db       6255 ATCGTTCCCGTCGAGTCGCAGGCGCTGGGCACCCCAGTCATCGCCTATGGCAAGGGCGGC 6314

Qy        916 GCCTTAGAAACCGTTCGCCCACTAGGTG---TAGAGGAACCGACTGGCATTTTCTTCAAG 972
              |   | ||||||||    ||||| || |   | |  | ||||| ||| | |  |||
Db       6315 GTACTGGAAACCGTCATTCCACTCGGCGCCTCAAACCAGCCGACCGGCCTCTATTTCGGC 6374

Qy        973 GAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAAAATGCATCATTT 1032
              || ||||      |  |  |  ||  || ||     ||||| |   | ||| |      |
Db       6375 GAGCAGACACCGGAGGCGATCTGTGCCGCGGTCCAGGAATTCGTCGACAATTCCGACCGT 6434

Qy       1033 TTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCAAGATTTGAACAA 1092
              ||          || ||     | ||||| |     || || ||      ||||
Db       6435 TTCGACAAGAGTGCCTGCGTCGACAATGCCGCCCGCTTCTCGCGCGAGCGATTCCTGAGG 6494

Qy       1093 GAATT 1097
              |||||
Db       6495 GAATT 6499
```

```
RESULT 11
AE008357/c
LOCUS       AE008357               11355 bp    DNA     linear   BCT 28-MAY-2004
DEFINITION  Agrobacterium tumefaciens str. C58 linear chromosome, section 161
            of 187 of the complete sequence.
ACCESSION   AE008357 AE007870
VERSION     AE008357.1  GI:15160112
KEYWORDS    .
SOURCE      Agrobacterium tumefaciens str. C58
  ORGANISM  Agrobacterium tumefaciens str. C58
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Agrobacterium.
REFERENCE   1  (bases 1 to 11355)
  AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
  TITLE     Complete Genome Sequence of Agrobacterium tumefaciens C58
            (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
            Disease in Plants
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11355)
  AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney
            Street, Cambridge, MA 02139, USA
COMMENT     Approximately 800 bp of telomeric sequence missing from the left
            end of the chromosome and 200 bp missing from the right end.
FEATURES             Location/Qualifiers
     source          1. .11355
                     /organism="Agrobacterium tumefaciens str. C58"
                     /mol_type="genomic DNA"
                     /strain="C58"
                     /isolate="Cereon"
                     /db_xref="taxon:176299"
     gene            complement(232. .1455)
                     /gene="AGR_L_3095"
     CDS             complement(232. .1455)
                     /gene="AGR_L_3095"
                     /note="hypothetical protein 18.9 - Salmonella
                     choleraesuis"
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_3095p"
                     /protein_id="AAK90119.1"
                     /db_xref="GI:15160113"
                     /translation="MFSRAADRPNIRYRRACSKRTINFYYVEIIEEIMKTAIVHDWLT
                     DRGGAEKVLHNLLEIYPHADLYCLVDFMSERDRGFLGGRPVNTSFIQKLPFARKKYRS
                     YLPFMPLAIEQFDLSGYDLVISSSYAVAKGVITGPGQFHVSYIHSPIRYAWDLQHQYL
                     KETRLNRGFASWIARAILHYIRMWDIRTTHGVDLMTVNSKFIRSRVRKCYGRDSTVIY
                     PPVDTAHLSPSDRKGDYFVTASRLVPYKKVHLIVEAFAKMPDKRLVVIGDGPDMKRIR
                     DIATPNVEILGFVGNEELHKYMAEAKAFVFAAEDDFGIVPVESQALGTPVIAYGKGGV
                     LETVIPLGASNQPTGLYFGEQTPEAICAAVQEFVDNSDRFDKSACVDNAARFSRERFL
                     REFADTVNFALAERA"
gene                 1493. .1735
                     /gene="AGR_L_3096"
CDS                  1493. .1735
                     /gene="AGR_L_3096"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_3096p"
                     /protein_id="AAK90120.1"
                     /db_xref="GI:15160114"
                     /translation="MGGYVCPSPIFNNTSFANAKQPSCVILRRRFPRECVQQYPPCRL
                     APFCSPEKSCFQFADTGNPSSRRKVIDKASYRDILL"
gene                 1768. .3321
                     /gene="AGR_L_3099"
CDS                  1768. .3321
                     /gene="AGR_L_3099"
                     /note="(AF065159) putative ABC transporter"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_3099p"
                     /protein_id="AAK90121.1"
                     /db_xref="GI:15160115"
                     /translation="MEQRAGSLTMRRKIAFLATAALLLAPGVMTAAEKGGVINVATIG
                     EPPTLDPMASTADLVGIVTQHIFETLYTFDKGWKVTPLLAESLPEISADGKTYTIKLR
                     QGIKFHDGSDMTSEDVVASLGRWMKIASRGKQAAGFIDNISAPDAGTVTITLKQPYAP
                     LTSLLAFNNSAAIILPAEKQAEPMAEFVGTGPYMLKERKADQYIQLVRFDGYKSRDGE
                     SDGYGGARHQYLDEIRFVPVPDPNTRVEAAVSGQYDYVDSIPVESFEKVKSSSASQPM
                     MLKPFGFPVFVFNTKEGLASKLEVRKAVTEALNMEDMLAAAFGSTDFYALDGAYYPDN
                     FSWHTEEGVDGNYNVGNPEAAAEKLKAAGYNGEPLRILTSRQYEFHYKMAQVAAEYLK
                     LAGFKVDMQVVDWATLTQRRADPKLWDIYISHSPFLPEPALMGALSTSSPGWWDTPAR
                     KAAVDAFTSEADPEKRLALWANVQKTIYAELPLIKIGDFNAAAAKSNKLEGVDPAPWP
                     YFWNASVTK"
gene                 3651. .4592
                     /gene="AGR_L_3101"
CDS                  3651. .4592
                     /gene="AGR_L_3101"
                     /note="PROBABLE PEPTIDE ABC TRANSPORTER PERMEASE PROTEIN
                     Y4TP"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_3101p"
                     /protein_id="AAK90122.1"
                     /db_xref="GI:15160116"
                     /translation="MIRYIFQRLAGMIVVMFLVVTIVFVILRVTPGDPAAVMLGPDAT
                     AQDIAELRTRLGLDQSLGVQYVYYIGQLLKGDLGQSIFLNMPVGSALLDRAEPTFFLT
                     IFSLLIAGVIALPIGIYAAYRRGSFVDQAATTIAMLAASIPSFWLGLILMQFFAVRLD
                     LFPVSGYGGPGASFLERMYHLTLPAIALGLVSSALILRFTRASMLDVLGDDYVRTARA
                     KGVKERRVVMKHALKNALIPILTVLGLTAAVLISGAVVTETVFGLPGVGNLVVSAVLR
                     RDYPVIQGALLVIAGLYVLINFAIDMLYLLVDPRVRY"
gene                 4592. .5467
                     /gene="AGR_L_3102"
CDS                  4592. .5467
                     /gene="AGR_L_3102"
                     /note="dipeptide abc transporter, dipeptide-binding
                     protein PAB0093 - Pyrococcus abyssi (strain Orsay)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_3102p"
                     /protein_id="AAK90123.1"
                     /db_xref="GI:15160117"
                     /translation="MADITSTAPPTRNPTILFIRRLLKRRTVAAGLIVLLVFVLLAVF
                     APMIAPYSPSKLSIVNRLKPPSELYWFGTDEFGRDVFSRTIYAGRLSLLVGAAVVALS
                     SIIGITLGLLAGFFQKLDTPIARLIDAMMAFPDILLAIALVAALGPSLTTVIIALAVV
                     YSPRLARIVRASTLVIRELPYVEAARALGISTFHIMTRHVLRNLVSPILVQCTFLFAS
                     AMLAEAGLSFLGLGVSPEIPTWGTMISAGRQYIGQADWMTYFPGFAIILSVLSLQMVG
                     DGLRDMLDPRLRRDL"
gene                 5379. .6692
                     /gene="AGR_L_3103"
CDS                  5379. .6692
                     /gene="AGR_L_3103"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_3103p"
                     /protein_id="AAK90124.1"
                     /db_xref="GI:15160118"
                     /translation="MPSFFPCCRCKWSATGCGTCSTQDCEGICKMTSGEQAKTPLQAP
                     ILLTNVKPVGFGKGASQSSTDILIGGDGKIAAVGSALQAPADTQRIDAKGAFISPGWV
                     DLHVHIWHGGTDISIRPSECGAERGVTTLVDAGSAGEANFHGFREYIIEPSRERIKAF
                     LNLGSIGLVACNRVPELRDIKDIDLDRILECYAENSEHIVGLKVRASHVITGSWGVTP
                     VKLGKKIAKILKVPMMVHVGEPPALYDEVLEILGPGDVVTHCFNGKSGSSIMEDEDLF
                     NLAERCAGEGIRLDIGHGGASFSFKVAEAAIARGLLPFSISTDLHGHSMNFPVWDLAT
                     TMSKLLSVDMPFENVVEAVTRNPASVIRLDMENRLDVGQRADFTVFDLVDADLEATDS
                     NGDVSRLKRLFEPRYAVIGAEAIAASRYIPRARKLVRHSHGYSWR"
gene                 6776. .7270
                     /gene="AGR_L_3105"
CDS                  6776. .7270
                     /gene="AGR_L_3105"
                     /note="probable transcription regulator Rv3678c -
                     Mycobacterium tuberculosis (strain H37RV)"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_3105p"
                     /protein_id="AAK90125.1"
                     /db_xref="GI:15160119"
                     /translation="MKQSLLDAEGHLSPYDRLAALGISLPPPPPPIANFVTHVVEGNM
                     LYLSGQGPREENGHLYAGKVGAEIGLEEAYNHARLTGINLLAVMHEALGDLSRVKKVV
                     KLLGMVNAVSDFRDHPSVINGCSDLFIDVFGDAGQHARSAVGFGSLPGNITVEIEAIV
                     ALKG"
gene                 complement(7277. .8128)
                     /gene="AGR_L_3107"
CDS                  complement(7277. .8128)
                     /gene="AGR_L_3107"
                     /note="(AF307355) putative repressor"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_3107p"
                     /protein_id="AAK90126.1"
                     /db_xref="GI:15160120"
                     /translation="MRSRNLDAHRLDEIASTALSDTAENPDVKAAKRSRVSGIDRALQ
                     VIDYLYETGSPAGAYAIAKAIKAPLSTVYVIVDDLVEKTMLARNADGNIWLGARLYHY
                     GLAYARSLDFMGVATHEMHDLCRHAGETVQLCGRDGDHMLVLAMADGPSHFQVASRVG
                     TRVPLNWTASGRLLVGHLPETERLELFRRCARSSPTGRAEINPETLSQSAAAAFHDRL
                     SIQVAESDYAVACIASPICDINGECVATISIVLPEQKVLADENHYTEQVRVSAERIEK
                     LMGWRNH"
gene                 8080. .9399
                     /gene="AGR_L_3109"
CDS                  8080. .9399
                     /gene="AGR_L_3109"
                     /note="(Y14814) selenocysteine synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="AGR_L_3109p"
                     /protein_id="AAK90127.1"
                     /db_xref="GI:15160121"
                     /translation="MKLFRPVDGRPNSGSSCILRSYYIIMAYMSMPIASQRQEKDMTE
                     DIRSRIGLRPVINVSGTMTSLGASIVVPEAVEAMAAILPQFVEINDLQRKASAIIARL
                     TGGEAGFVTASCSSGISLAVAGAITGNNLLAIEKLPDIAPEKNEVLVQMGHVVSYGAP
                     VDQAIRLGGGKVVLVGQATSTHRYHMENAITEKTAAAVYVVSHHVVDYGLLHLSEFVE
                     IAHAKGVPVIVDAASEYDLELFLATGADVVLYSGHKFLGGPTSGIVAGSKELVRHAFL
                     QNMGIGRGMKVGKESIYGVMAALEAWEKRDHAGIRERETGYLNLWKKTLDGRPGVTAL
                     IEPDPTNNPLDRMRVIIDADEAHITAWDLTTALARGNPPIITRDHEVEHRYFYLDPCN
                     LHPGQETIVASRLAEELDKARASNEMIATPFEDRSRHRFDGMLCWPD"
```

```
    gene            9455. .11290
                    /gene="AGR_L_3112"
    CDS             9455. .11290
                    /gene="AGR_L_3112"
                    /note="(D90720) Hypothetical ABC transporter ATP-binding
                    protein in bcr 5'region."
                    /codon_start=1
                    /transl_table=11
                    /product="AGR_L_3112p"
                    /protein_id="AAK90128.1"
                    /db_xref="GI:15160122"
                    /translation="MTMPQAAPLGAEDAVLSVRNLTTSFHVDGGWKPVVRDISFDVGP
                    GETVAIVGESGSGKSVTSLSIMRLLDRESSRVQGEILLGGRNLLSLSEDAMRRVRGNE
                    VAMIFQEPMTSLNPLFTIGDQISEALLCHQPMSKAEARAETVRLLEKVRIPSAASRFD
                    EYPHRFSGGMRQRVMIAMALASKPKLLIADEPTTALDVTIQGQILDLIKTLQEEEGTS
                    VLFITHDMGVVAEIADRTVVMYRGEQVEVGATADIFHRGKHPYTRALLSAVPVLGSMK
                    GEERPLRFPIVNTATGETQEPVRPADTVDASAQPVLKVEGLTKRFDIHSGLLGRLSGR
                    VHAVENVSFDLRAGETLSLVGESGCGKSTTGRAIMRLIEPDAGSVVVNGENILTLDKS

  Query Match             23.2%;  Score 269;  DB 1;  Length 11355;
  Best Local Similarity   53.6%;  Pred. No. 3.5e-43;
  Matches  582;  Conservative    0;  Mismatches  500;  Indels    3;  Gaps    1;

Qy         16 ATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTCCGAACAGGTATCA 75
Db       1356 ATGAAGACGGCAATCGTGCATGACTGGCTCACCGACCGCGGTGGGGCCGAGAAAGTGCTT 1297

Qy         76 TCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGTTGATTTTCTAACG 135
Db       1296 CACAACCTTCTCGAAATTTATCCTCACGCGGACCTTTATTGCCTCGTCGATTTCATGAGC 1237

Qy        136 GATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATTTATTCAAAATTTA 195
Db       1236 GAGCGCGACCGGGGCTTTCTCGGCGGGCGTCCCGTCAACACCTCCTTCATCCAGAAGCTG 1177

Qy        196 CCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCTATTGAACAA 255
Db       1176 CCCTTCGCCCGGAAGAAATATCGCTCCTATCTGCCGTTTATGCCGCTCGCCATCGAGCAG 1117

Qy        256 CTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCAAAAGGTGTT 315
Db       1116 TTCGATCTTAGCGGTTATGACCTCGTCATCTCCTCCAGCTATGCGGTCGCCAAGGGCGTC 1057

Qy        316 ATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTATTCGATATGCGTGG 375
Db       1056 ATCACCGGTCCCGGCCAGTTTCACGTCTCCTACATCCACAGCCCCATTCGCTACGCCTGG 997

Qy        376 GATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAAGGTTGGTTA 435
Db        996 GATCTCCAGCATCAATATCTCAAGGAAACCCGGCTCAATCGCGGCTTCGCCTCGTGGATT 937

Qy        436 GCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAATGGGGTTGAT 495
Db        936 GCGCGCGCCATTCTTCACTACATTCGCATGTGGGATATCCGCACCACCCATGGCGTCGAT 877

Qy        496 CATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAG 555
Db        876 CTGATGACGGTGAATTCCAAATTCATCCGCAGCCGCGTGCGCAAATGTTATGGGCGCGAT 817

Qy        556 GCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAA 615
Db        816 TCCACGGTGATCTACCCGCCGGTCGATACCGCGCATCTTTCCCCTTCCGACAGGAAAGGC 757

Qy        616 GACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAA 675
Db        756 GATTATTTCGTCACCGCCTCACGGCTGGTGCCCTACAAGAAAGTGCACCTGATCGTGGAG 697

Qy        676 GCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAA 735
Db        696 GCCTTTGCGAAGATGCCCGATAAACGGCTTGTCGTGATCGGCGACGGGCCGGACATGAAA 637

Qy        736 AAAATAAAGAGCAAGGCTACAGACAATATATAAAATTGCTCGGTTATCAATCTTTTCCTGTT 795
Db        636 CGCATCCGCGACATCGCCACGCCTAATGTCGAAATTCTCGGTTTTGTCGGCAATGAGGAG 577

Qy        796 TTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGA 855
Db        576 CTTCACAAATACATGGCGGAAGCCAAGGCATTCGTTTTTGCCGCTGAAGATGATTTCGGC 517

Qy        856 ATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGG 915
Db        516 ATCGTTCCCGTCGAGTCGCAGGCGCTGGGCACCCCAGTCATCGCCTATGGCAAGGGCGGC 457

Qy        916 GCCTTAGAAACCGTTCGCCCACTAGGTG---TAGAGGAACCGACTGGCATTTTCTTCAAG 972
Db        456 GTACTGGAAACCGTCATTCCACTCGGCGCCTCAAACCAGCCGACCGGCCTCTATTTCGGC 397

Qy        973 GAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAAAATGCATCATTT 1032
Db        396 GAGCAGACACCGGAGGCGATCTGTGCCGCGGTCCAGGAATTCGTCGACAATTCCGACCGT 337

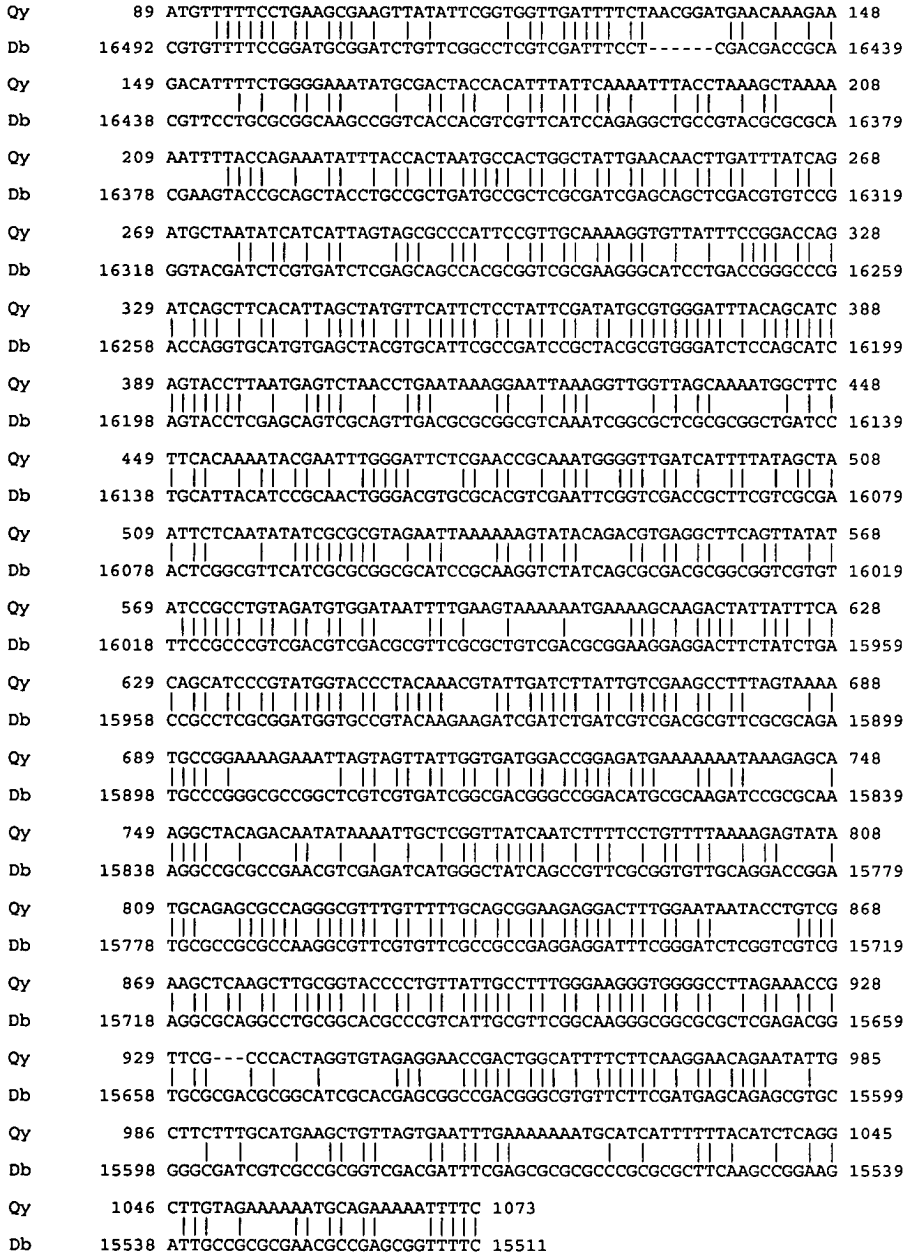
Qy       1033 TTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCAAGATTTGAACAA 1092
Db        336 TTCGACAAGAGTGCCTGCGTCGACAATGCCGCCCGCTTCTCGCGCGAGCGATTCCTGAGG 277

Qy       1093 GAATT 1097
Db        276 GAATT 272


RESULT 12
BX571966_23/c
WPCOMMENT
  Sequence split into 32 fragments  LOCUS BX571966 Accession BX571966
     Fragment Name      Begin     End
     BX571966_00            1    110000
     BX571966_01       100001    210000
     BX571966_02       200001    310000
     BX571966_03       300001    410000
     BX571966_04       400001    510000
     BX571966_05       500001    610000
     BX571966_06       600001    710000
     BX571966_07       700001    810000
     BX571966_08       800001    910000
     BX571966_09       900001   1010000
     BX571966_10      1000001   1110000
     BX571966_11      1100001   1210000
     BX571966_12      1200001   1310000
     BX571966_13      1300001   1410000
     BX571966_14      1400001   1510000
     BX571966_15      1500001   1610000
     BX571966_16      1600001   1710000
     BX571966_17      1700001   1810000
     BX571966_18      1800001   1910000
     BX571966_19      1900001   2010000
     BX571966_20      2000001   2110000
     BX571966_21      2100001   2210000
     BX571966_22      2200001   2310000
     BX571966_23      2300001   2410000
     BX571966_24      2400001   2510000
     BX571966_25      2500001   2610000
     BX571966_26      2600001   2710000
     BX571966_27      2700001   2810000
     BX571966_28      2800001   2910000
     BX571966_29      2900001   3010000
     BX571966_30      3000001   3110000
     BX571966_31      3100001   3173005
Continuation (24 of 32) of BX571966 from base 2300001 (BX571966 Burkholderia pseudomallei

  Query Match             23.1%;  Score 267.6;  DB 1;  Length 110000;
  Best Local Similarity   54.9%;  Pred. No. 3.6e-43;
  Matches  575;  Conservative    0;  Mismatches  464;  Indels    9;  Gaps    2;

Qy         29 TCGTTCATGAGTGGTTATTAAGTTATGCAGGCTCCGAACAGGTATCATCTGCCATCCTGC 88
Db      16552 TCGTGCACGACTGGCTCGTCACCTACGCGGGCGCCGAGCGCGTGCTCGAGCAGATCGTCG 16493
```



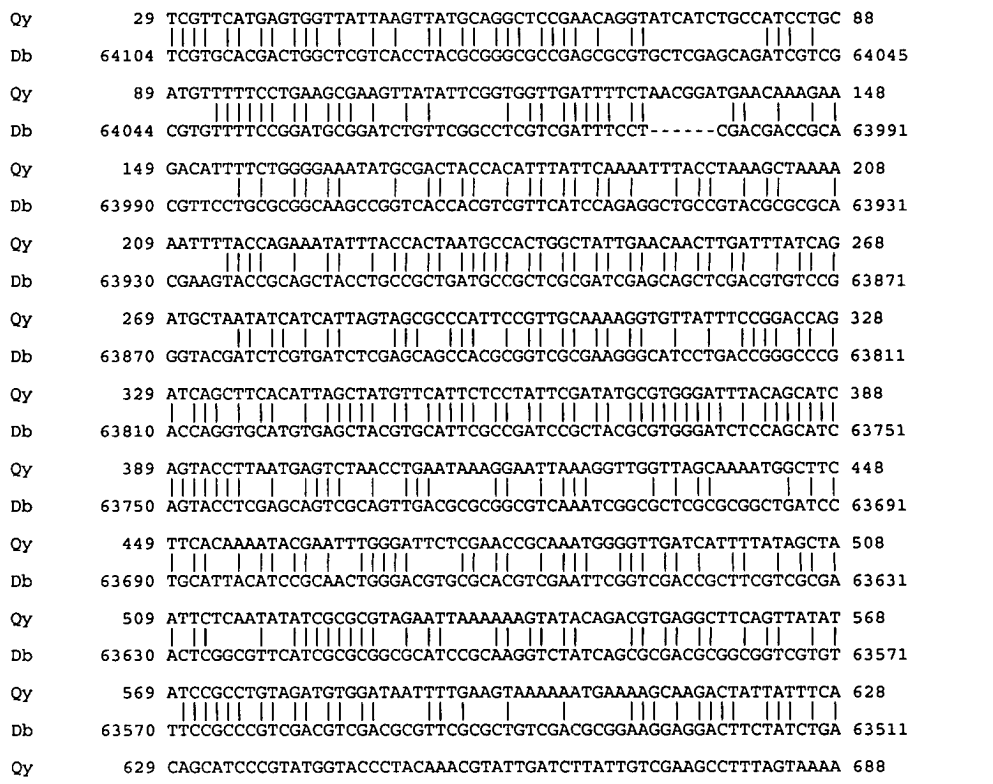
RESULT 13
CP000011_18/c
WPCOMMENT

Sequence split into 24 fragments LOCUS CP000011 Accession CP000011

Fragment Name	Begin	End
CP000011_00	1	110000
CP000011_01	100001	210000
CP000011_02	200001	310000
CP000011_03	300001	410000
CP000011_04	400001	510000
CP000011_05	500001	610000
CP000011_06	600001	710000
CP000011_07	700001	810000
CP000011_08	800001	910000
CP000011_09	900001	1010000
CP000011_10	1000001	1110000
CP000011_11	1100001	1210000
CP000011_12	1200001	1310000
CP000011_13	1300001	1410000
CP000011_14	1400001	1510000
CP000011_15	1500001	1610000
CP000011_16	1600001	1710000
CP000011_17	1700001	1810000
CP000011_18	1800001	1910000
CP000011_19	1900001	2010000
CP000011_20	2000001	2110000
CP000011_21	2100001	2210000
CP000011_22	2200001	2310000
CP000011_23	2300001	2325379

Continuation (19 of 24) of CP000011 from base 1800001 (CP000011 Burkholderia mallei ATCC

Query Match 23.1%; Score 267.6; DB 1; Length 110000;
Best Local Similarity 54.9%; Pred. No. 3.6e-43;
Matches 575; Conservative 0; Mismatches 464; Indels 9; Gaps 2;



```
                 | || || || |||||  || |||||     || ||||| || ||||| || ||      | |
Db      63510 CCGCCTCGCGGATGGTGCCGTACAAGAAGATCGATCTGATCGTCGACGCGTTCGCGCAGA 63451

Qy        689 TGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAAAAAATAAAGAGCA 748
              |||| |           | || || || || || || ||||| |||    || ||       |
Db      63450 TGCCCGGGCGCCGGCTCGTCGTGATCGGCGACGGGCCGGACATGCGCAAGATCCGCGCAA 63391

Qy        749 AGGCTACAGACAATATAAAATTGCTCGGTTATCAATCTTTTCCTGTTTTAAAAGAGTATA 808
              ||||  |    ||  |  |  |  | || |||||  | ||  | || ||  | ||    |
Db      63390 AGGCCGCGCCGAACGTCGAGATCATGGGCTATCAGCCGTTCGCGGTGTTGCAGGACCGGA 63331

Qy        809 TGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCG 868
              |||   |||||| |||||| || || || || || ||||| || || ||      ||||
Db      63330 TGCGCCGCGCCAAGGCGTTCGTGTTCGCCGCCGAGGAGGATTTCGGGATCTCGGTCGTCG 63271

Qy        869 AAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCG 928
              | || || || ||||| || || || ||||| || || ||||| || ||  | || || |
Db      63270 AGGCGCAGGCCTGCGGCACGCCCGTCATTGCGTTCGGCAAGGGCGGCGCGCTCGAGACGG 63211

Qy        929 TTCG---CCCACTAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTG 985
              | ||   | |   |      |||   ||||| ||| | |||||| | || ||||   |
Db      63210 TGCGCGACGCGGCATCGCACGAGCGGCCGACGGGCGTGTTCTTCGATGAGCAGAGCGTGC 63151

Qy        986 CTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGG 1045
                |   |   | || ||    || || ||       |  |   || |  |  | | |
Db      63150 GGGCGATCGTCGCCGCGGTCGACGATTTCGAGCGCGCGCCCGCGCGCTTCAAGCCGGAAG 63091

Qy       1046 CTTGTAGAAAAAATGCAGAAAAATTTTC 1073
                |||  |    || || ||    |||||
Db      63090 ATTGCCGCGCGAACGCCGAGCGGTTTTC 63063
```

RESULT 14
CP000010_00
WPCOMMENT
Sequence split into 36 fragments LOCUS CP000010 Accession CP000010

Fragment Name	Begin	End
CP000010_00	1	110000
CP000010_01	100001	210000
CP000010_02	200001	310000
CP000010_03	300001	410000
CP000010_04	400001	510000
CP000010_05	500001	610000
CP000010_06	600001	710000
CP000010_07	700001	810000
CP000010_08	800001	910000
CP000010_09	900001	1010000
CP000010_10	1000001	1110000
CP000010_11	1100001	1210000
CP000010_12	1200001	1310000
CP000010_13	1300001	1410000
CP000010_14	1400001	1510000
CP000010_15	1500001	1610000
CP000010_16	1600001	1710000
CP000010_17	1700001	1810000
CP000010_18	1800001	1910000
CP000010_19	1900001	2010000
CP000010_20	2000001	2110000
CP000010_21	2100001	2210000
CP000010_22	2200001	2310000
CP000010_23	2300001	2410000
CP000010_24	2400001	2510000
CP000010_25	2500001	2610000
CP000010_26	2600001	2710000
CP000010_27	2700001	2810000
CP000010_28	2800001	2910000
CP000010_29	2900001	3010000
CP000010_30	3000001	3110000
CP000010_31	3100001	3210000
CP000010_32	3200001	3310000
CP000010_33	3300001	3410000
CP000010_34	3400001	3510000
CP000010_35	3500001	3510148

```
LOCUS       CP000010             3510148 bp    DNA     circular BCT 22-SEP-2004
DEFINITION  Burkholderia mallei ATCC 23344 chromosome 1, complete sequence.
ACCESSION   CP000010
VERSION     CP000010.1  GI:52426793
KEYWORDS    .
SOURCE      Burkholderia mallei ATCC 23344
  ORGANISM  Burkholderia mallei ATCC 23344
            Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
            Burkholderiaceae; Burkholderia.
REFERENCE   1  (bases 1 to 3510148)
  AUTHORS   Nierman,W.C., DeShazer,D., Kim,H.S., Tettelin,H., Nelson,K.E.,
            Feldblyum,T., Ulrich,R.L., Ronning,C.M., Brinkac,L.M.,
            Daugherty,S.C., Davidson,T.D., Deboy,R.T., Dimitrov,G.,
            Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Khouri,H.,
            Kolonay,J.F., Madupu,R., Mohammoud,Y., Nelson,W.C., Radune,D.,
            Romero,C.M., Sarria,S., Selengut,J., Shamblin,C., Sullivan,S.A.,
            White,O., Yu,Y., Zafar,N., Zhou,L. and Fraser,C.M.
  TITLE     Structural flexibility in the Burkholderia mallei genome
  JOURNAL   (er) Proc. Natl. Acad. Sci. U.S.A. 101 (39), 14247-14251 (2004)
REFERENCE   2  (bases 1 to 3510148)
  AUTHORS   Nierman,W.C., DeShazer,D., Kim,H.S., Tettelin,H., Nelson,K.E.,
            Feldblyum,T., Ulrich,R.L., Ronning,C.M., Brinkac,L.M.,
            Daugherty,S.C., Davidson,T.D., Deboy,R.T., Dimitrov,G.,
            Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Khouri,H.,
            Kolonay,J.F., Madupu,R., Mohammoud,Y., Nelson,W.C., Radune,D.,
            Romero,C.M., Sarria,S., Selengut,J., Shamblin,C., Sullivan,S.A.,
            White,O., Yu,Y., Zafar,N., Zhou,L. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2004) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .3510148
                     /organism="Burkholderia mallei ATCC 23344"
                     /mol_type="genomic DNA"
                     /strain="ATCC 23344"
                     /db_xref="ATCC:23344"
                     /db_xref="taxon:243160"
                     /chromosome="1"
     gene            70. .1626
                     /gene="dnaA"
                     /locus_tag="BMA0001"
     CDS             70. .1626
                     /gene="dnaA"
                     /locus_tag="BMA0001"
                     /note="identified by match to protein family HMM PF00308;
                     match to protein family HMM TIGR00362"
                     /codon_start=1
                     /transl_table=11
                     /product="chromosomal replication initiator protein DnaA"
                     /protein_id="AAU48917.1"
                     /db_xref="GI:52428324"
                     /translation="MTPQQYVTWIKPLAPVAFDAAANTLSIAAPNRFKLDWVKSQFSG
                     RISDLARDFWNAPIEVQFVLDPKAGQRSPAGATPLAPRAPLPSANPAPVAPGPACAPA
                     VDAHAPAPAGMNAATAAAVAAAQAAQAAQANAAALNADEAADLDLPSLTAHEAAAGRR
                     TWRPGAANANSEAADSMYERSKLNPVLTFDNFVTGKANQLARAAAIQVADNPGISYNP
                     LFLYGGVGLGKTHLIHAIGNQLLLDKPGARIRYIHAEQYVSDVVKAYQRKAFDDFKRY
                     YHSLDLLLIDDIQFFSGKSRTQEEFFYAFEALVANKAQVIITSDTYPKEISGIDDRLI
                     SRFDSGLTVAIEPPELEMRVAILMRKAQSEGVSLSEDVAFFVAKHLRSNVRELEGALR
                     KILAYSKFHGREITIELTKEALKDLLTVQNRQISVENIQKTVADFYNIKVADMYSKKR
                     PANIARPRQIAMYLAKELTQKSLPEIGELFGGRDHTTVLHAVRKIADERGKDAQLNHE
                     LHVLEQTLKG"
     repeat_region   806. .813
                     /note="simple sequence repeat 2mer-1"
                     /rpt_family="SSR 2mer"
     gene            1789. .2892
                     /gene="dnaN"
                     /locus_tag="BMA0002"
     CDS             1789. .2892
                     /gene="dnaN"
                     /locus_tag="BMA0002"
                     /EC_number="2.7.7.7"
                     /note="identified by match to protein family HMM PF00712;
                     match to protein family HMM PF02767; match to protein
```

```
                family HMM PF02768; match to protein family HMM TIGR00663"
                /codon_start=1
                /transl_table=11
                /product="DNA polymerase III, beta subunit"
                /protein_id="AAU48918.1"
                /db_xref="GI:52428325"
                /translation="MQLVKTERDTLLRPLQTVSGIVERRHTLPILANLLITKNGPDVS
                FLSTDLELQITTRADFGVGGDQVATTVAARKLLDILRAMPDGQVTLSLADKRLTVQSG
                KSRFALQTLAADEFPTVAQSKDFGASLAVPQKAFRQLLGMVYFAMAQQDIRYYLNGML
                LVVDGDRLMAVATDGHRLAFSSMKIEGSFGRQEVIVPRKTILELQRLLEDIDDTVKID
                IAQTQAKFTFGQVELVSKLVEGKFPDFQRVIPKAHKNTFEIGREELQRSLQRAAILTS
                DKFKGVRCIIAPGQLKIMSTNADQEEAQEELEIAYQGDTVDIGFNVTYLLDVLANLKV
                DTVQVSLGDASSSALITVPENDEFKYVVMPMRI"
repeat_region   2893. .2900
                /note="simple sequence repeat 2mer-2"
                /rpt_family="SSR 2mer"
repeat_region   3021. .3026
                /note="simple sequence repeat 1mer-1"
                /rpt_family="SSR 1mer"
gene            3081. .5549
                /gene="gyrB"
                /locus_tag="BMA0003"
CDS             3081. .5549
                /gene="gyrB"
                /locus_tag="BMA0003"
                /EC_number="5.99.1.3"
                /note="identified by match to protein family HMM PF00204;
                match to protein family HMM PF00986; match to protein
                family HMM PF01751; match to protein family HMM PF02518;
                match to protein family HMM TIGR01059"
                /codon_start=1
                /transl_table=11
                /product="DNA gyrase, B subunit"
                /protein_id="AAU48919.1"
                /db_xref="GI:52428326"
                /translation="MTEQHNSQPENSYGASSIQILEGLEAVRKRPGMYIGDTSDGTGL
                HHLVFEVLDNSIDEALAGYCNDIHVTIHADNSISVTDNGRGIPTDVKLNDKHEPKRSA
                AEIVMTELHAGGKFDQNSYKVSGGLHGVGVSCVNALSSWLRLIVRRDGKKHFMEFHRG
                IAQNRVLEERDGEQVSPMQVIGETENRGTEVHFMADPTIFGTVEYHYDILAKRIRELS
                FLNNGVRIRLTDLRSGKEDDFAFAGGVKGFVEYINKTKSVLHPTIFHINGEKDSVGVE
                VAMQWNDSYNENVLCFTNNIPQRDGGTHLTGLRAAMTRVINKYIADNEIAKKAKVETT
                GDDMREGLSCVLSVKVPEPKFSSQTKDKLVSSEVRAPVEEVVAKALEEFLLETPNDAK
                IICGKIVEAARARDAARKAREMTRRKGVLDGVGLPGKLADCQEKDPAKSEIYIVEGDS
                AGGSAKQGRDRKFQAILPLRGKVLNVEKARYDKLLSSEQIVTLVTALGCGIGKDDYNL
                DKLRYHRIIIMTDADVDGAHIRTLLLTFFYRQMPEMIERGYVYIAQPPLYKVKAGRDE
                RYLKDDTELNAHMLRLALQGSELVPTENGTPISGDALGELARSYLLARGVVERLSRLY
                DPAALEAVMDGVAIDLSSEASTEASAKALAAMLSDDSKTEVRVVPAYDAVREQRSLRV
                ERTHHGNVRVSVIDEEFQHTADYQQLVNTANTFKGLIGAGAAIKRGERSSNVTDFKEA
                MKWLMADAERNLSKQRYKGLGEMNPEQLWETTMDPAVRRLLRVQIEDAIAADGIFTTL
                MGDEVEPRRAFIENNALRAGNIDV"
repeat_region   4184. .4191
                /note="simple sequence repeat 2mer-3"
                /rpt_family="SSR 2mer"
repeat_region   4284. .4296
                /note="simple sequence repeat 2mer-4"
                /rpt_family="SSR 2mer"
repeat_region   6007. .7242
                /locus_tag="BMA_IS407A-1"
                /rpt_family="IS407A"
                /insertion_seq="IS407A"
gene            6075. .6338
                /locus_tag="BMA0004"
CDS             6075. .6338
                /locus_tag="BMA0004"
                /note="identified by similarity to SP:P24580; match to
                protein family HMM PF01527"
                /codon_start=1
                /transl_table=11
                /product="IS407A, transposase OrfA"
                /protein_id="AAU50344.1"
                /db_xref="GI:52429751"
                /translation="MKKRFTEQQIIGFLKEAEAGMPVKELCRKHGFSDASFYTWRAKF
                GGMEVSEARRLKGLEVENARLKKLLAEAMLDMEALKVVVKGKP"
gene            6362. .7195
                /locus_tag="BMA0005"
CDS             6362. .7195
                /locus_tag="BMA0005"
                /note="identified by similarity to GP:17427840; match to
                protein family HMM PF00665"
                /codon_start=1
                /transl_table=11
                /product="IS407A, transposase OrfB"
                /protein_id="AAU48920.1"
                /db_xref="GI:52428327"
                /translation="MLAIREKVNISERRACRLVGLSRSVLHYDAKPDHENEVLAARLV
                KLAHERRRFGYRRLHALVEREGTHANHKRIYRLYREAGLAVRRRRKRHGVMIEREQLA
                LPGAPNEVWSIDFVMDALSNGRRVKCLTVVDDFTKEAVDIVVDHGISGLYVARALDRA
                ARFRGYPKAVRTDQGPEFTSRALDQWAYANGVTLKLIQAGKPTQNAYIESFNGKFRDE

  Query Match             22.2%;  Score 257.4;  DB 1;  Length 110000;
  Best Local Similarity   55.3%;  Pred. No. 3.7e-41;
  Matches  545;  Conservative    0;  Mismatches  431;  Indels    9;  Gaps    2;

Qy         17 TAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTCCGAACAGGTATCAT 76
Db      38156 TGAAGGTAGCAATCGTTCACGACTGGCTGGTGGTGTATGGCGGCGCGGAGCGTGTGCTCG 38215

Qy         77 CTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGTTGATTTTCTAACGG 136
Db      38216 CGCAGATGATCGACTGCTTTGCGCAAGCCGACATCTACAGCCTCGTCGATTTTCT----- 38270

Qy        137 ATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATTTATTCAAAATTTAC 196
Db      38271 -CGACGACCGCTCGTGCCTGCGTGGCCGGCCGGTGCACACCTCGTTCATCCAGAAATTGC 38329

Qy        197 CTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCTATTGAACAAC 256
Db      38330 CGTTCGCGCGCAGCAAGTACCGCAGCTATTTGCCGCTCTTTCCGCTCGCGATCGAGCAGT 38389

Qy        257 TTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCAAAAGGTGTTA 316
Db      38390 TCGATCTGTCCGGCTACGACCTGATCCTGTCGAGCTCGTATGCGGTCGCCAAGGGCGTGC 38449

Qy        317 TTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTATTCGATATGCGTGGG 376
Db      38450 TGAACGGCCCGGACCAGTTGCATGCGAGTTACGTGCACTCGCCCGTGCGCTACGCGTGGG 38509

Qy        377 ATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAAGGTTGGTTAG 436
Db      38510 ACCTGCAGCATCAGTACCTGAACGAAGCGGGGCTCGCGCGCGGCGTGAAATCGGCGCTCG 38569

Qy        437 CAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAATGGGGTTGATC 496
Db      38570 CGCGCACGTTGCTGCACTACATCCGCAACTGGGATGCGCGCTCGGCGAACGGGGTCGACC 38629

Qy        497 ATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGG 556
Db      38630 TGCTCGCGGCGAATTCGCGCTTCGTCGCGCGACGTATCCGCAAGACGTATCGGCGCGACG 38689

Qy        557 CTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAG 616
Db      38690 CGACGGTCATCTATCCGCCCGTCGACGTCGATCATCTCGCGCTGCGCGACACGAAGGACG 38749

Qy        617 ACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAAG 676
Db      38750 ACTTCTATCTGACGGCGTCGCGCCTCGTGCCGTACAAGCGGATCGATCTGATCGTCGAGG 38809

Qy        677 CCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAAA 736
Db      38810 CGTTTTCGCACATGCCGTCGCGCCGGCTCGTCGTGATCGGCGACGGGCCGGAGGCGGCGA 38869

Qy        737 AAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTATCAATCTTTTCCTGTTT 796
Db      38870 AGATCCGCGCGCTCGCGGGCCCGAACGTCACGCTGCTCGGCTACCAGCCGTTCGACGTGC 38929

Qy        797 TAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAA 856
```

```
Db      38930 TGCACGATCATCTGCAGCGCGCGAAGGCGTTCGTGTTCGCCGCGGAAGAGGATTTCGGCA 38989

Qy        857 TAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGG 916

Db      38990 TCTCGCCCGTCGAAGCGCAGGCATGCGGCACGCCCGTGATCGCATACGGCAAGGGCGGCG 39049

Qy        917 CCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAAC 976

Db      39050 TGTGCGAATCGGTGCGCGC---GGCGGGCGCGGCGCCGACGGGCCTCTTCTATGCGAAGC 39106

Qy        977 AGAATATTGCTTCTTTGCATGAAGC 1001

Db      39107 AAACGTGCGACGCGTTGATCGATGC 39131


RESULT 15
BX571965_06
WPCOMMENT
  Sequence split into 41 fragments  LOCUS BX571965 Accession BX571965
     Fragment Name      Begin     End
     BX571965_00            1    110000
     BX571965_01       100001    210000
     BX571965_02       200001    310000
     BX571965_03       300001    410000
     BX571965_04       400001    510000
     BX571965_05       500001    610000
     BX571965_06       600001    710000
     BX571965_07       700001    810000
     BX571965_08       800001    910000
     BX571965_09       900001   1010000
     BX571965_10      1000001   1110000
     BX571965_11      1100001   1210000
     BX571965_12      1200001   1310000
     BX571965_13      1300001   1410000
     BX571965_14      1400001   1510000
     BX571965_15      1500001   1610000
     BX571965_16      1600001   1710000
     BX571965_17      1700001   1810000
     BX571965_18      1800001   1910000
     BX571965_19      1900001   2010000
     BX571965_20      2000001   2110000
     BX571965_21      2100001   2210000
     BX571965_22      2200001   2310000
     BX571965_23      2300001   2410000
     BX571965_24      2400001   2510000
     BX571965_25      2500001   2610000
     BX571965_26      2600001   2710000
     BX571965_27      2700001   2810000
     BX571965_28      2800001   2910000
     BX571965_29      2900001   3010000
     BX571965_30      3000001   3110000
     BX571965_31      3100001   3210000
     BX571965_32      3200001   3310000
     BX571965_33      3300001   3410000
     BX571965_34      3400001   3510000
     BX571965_35      3500001   3610000
     BX571965_36      3600001   3710000
     BX571965_37      3700001   3810000
     BX571965_38      3800001   3910000
     BX571965_39      3900001   4010000
     BX571965_40      4000001   4074542
Continuation (7 of 41) of BX571965 from base 600001 (BX571965 Burkholderia pseudomallei

  Query Match             22.2%;  Score 257.2;  DB 1;  Length 110000;
  Best Local Similarity   55.6%;  Pred. No. 4.1e-41;
  Matches  518;  Conservative    0;  Mismatches  408;  Indels    6;  Gaps    1;

Qy         17 TAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTCCGAACAGGTATCAT 76

Db      84417 TGAAGGTAGCAATCGTTCACGACTGGCTGGTGGTGTATGGCGGCGCGGAGCGTGTGCTCG 84476

Qy         77 CTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGTTGATTTTCTAACGG 136

Db      84477 CGCAGATGATCGACTGCTTTGCGCAAGCCGACATCTACAGCCTCGTCGATTTCCT----- 84531

Qy        137 ATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATTTATTCAAAATTTAC 196

Db      84532 -CGACGACCGCTCGTGCCTGCGTGGCCGGCCGGTGCACACCTCGTTCATCCAGAAATTGC 84590

Qy        197 CTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCTATTGAACAAC 256

Db      84591 CGTTCGCGCGCAGCAAGTACCGCAGCTATTTGCCGCTCTTTCCGCTCGCGATCGAGCAGT 84650

Qy        257 TTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCAAAAGGTGTTA 316

Db      84651 TCGATCTGTCCGGCTACGACCTGATCCTGTCGAGCTCGTATGCGGTCGCCAAGGGCGTGC 84710

Qy        317 TTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTATTCGATATGCGTGGG 376

Db      84711 TGAACGGCCCGGACCAGTTGCATGCGAGTTACGTGCACTCGCCCGTGCGCTACGCGTGGG 84770

Qy        377 ATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAAGGTTGGTTAG 436

Db      84771 ACCTGCAGCATCAGTACCTGAACGAAGCGGGGCTCGCGCGCGGCGTGAAATCGGCGCTCG 84830

Qy        437 CAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAATGGGGTTGATC 496

Db      84831 CGCGCACGTTGCTGCACTACATCCGCAACTGGGATGCGCGCTCGGCGAACGGGGTCGACC 84890

Qy        497 ATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGG 556

Db      84891 TGCTCGCGGCGAATTCGCGCTTCGTCGCGCGACGTATCCGCAAGACGTATCGGCGCGACG 84950

Qy        557 CTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAG 616

Db      84951 CGACGGTCATCTATCCGCCCGTCGACGTCGATCATCTCGCGCTGCGCGACACGAAGGACG 85010

Qy        617 ACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAAG 676

Db      85011 ACTTCTATCTGACGGCGTCGCGCCTCGTGCCGTACAAGCGGATCGATCTGATCGTCGAGG 85070

Qy        677 CCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAAA 736

Db      85071 CGTTTTCGCACATGCCGTCGCGCCGGCTCGTCGTGATCGGCGACGGGCCGGAGGCGGCGA 85130

Qy        737 AAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTATCAATCTTTTCCTGTTT 796

Db      85131 AGATCCGCGCGCTCGCGGGCCCGAACGTCACGCTGCTCGGCTACCAGCCGTTCGACGTGC 85190

Qy        797 TAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAA 856

Db      85191 TGCACGATCATCTGCAGCGCGCGAAGGCGTTCGTGTTTGCCGCGGAAGAGGATTTCGGCA 85250

Qy        857 TAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGG 916

Db      85251 TCTCGCCCGTCGAAGCGCAGGCATGCGGCACGCCCGTGATCGCATACGGCAAGGGCGGCG 85310

Qy        917 CCTTAGAAACCGTTCGCCCACTAGGTGTAGAG 948

Db      85311 TGTGCGAATCGGTGCGCGCGGCGGGCGCGGCG 85342


Search completed: May 11, 2006, 19:14:44
Job time : 4176.16 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 07:09:59 ; Search time 527.645 Seconds
(without alignments)
14626.705 Million cell updates/sec

Title: US-09-423-093B-3_COPY_7310_8467
Perfect score: 1158
Sequence: 1 atggaaataaatgaaataaa.....aacagattattaaacgttaa 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1158	100.0	12441	2	AAX06750	Aax06750 Salmonell
	2	520	44.9	1167	11	ACH95204	Ach95204 Klebsiell
c	3	181.4	15.7	324	11	ACH95410	Ach95410 Klebsiell
c	4	120.2	10.4	222	11	ACH95331	Ach95331 Klebsiell
	5	54	4.7	2000	11	ACL35887	Acl35887 Rice stre
c	6	50.2	4.3	2000	11	ACL37108	Acl37108 Rice stre
c	7	47	4.1	2000	11	ACL35363	Acl35363 Rice stre
	8	45	3.9	2000	11	ACL37108	Acl37108 Rice stre
c	9	44	3.8	2000	11	ACL35887	Acl35887 Rice stre
	10	42.6	3.7	4590	5	AAH24065	Aah24065 Yeast AOD
c	11	42.6	3.7	110000	2	AAX20248_03	Continuation (4 of
	12	42.4	3.7	1756	12	ADG47988	Adg47988 Human UDP
	13	42.4	3.7	1766	5	AAS91380	Aas91380 DNA encod
	14	42.4	3.7	2406	5	AAS91382	Aas91382 DNA encod
	15	42.4	3.7	2491	9	AAD57521	Aad57521 Human enz
c	16	42.4	3.7	117813	14	ADX07540	Adx07540 Cyclin-de
	17	42.2	3.6	456	14	ACL65390	Acl65390 M. xanthu
	18	42.2	3.6	2000	6	ABZ16337	Abz16337 Arabidops
	19	42.2	3.6	2000	8	ADA69071	Ada69071 Arabidops
	20	42.2	3.6	2002	7	ADZ74514	Adz74514 Arabidops
c	21	42.2	3.6	5752	14	ACL64283	Acl64283 M. xanthu
	22	42.2	3.6	11155	6	ABL32604	Abl32604 Human imm
	23	41.2	3.6	6000	6	ABL34304	Abl34304 Human imm
	24	41.2	3.6	6621	3	AAA70188	Aaa70188 Plasmodiu
	25	41	3.5	1889	9	ADA10965	Ada10965 Human cDN
c	26	41	3.5	139326	13	ABD33155	Abd33155 Human can
c	27	40.4	3.5	2000	8	ADA71938	Ada71938 Rice gene
	28	40.2	3.5	6136	6	ABK40042	Abk40042 Human che
c	29	40.2	3.5	19659	6	ABL32767	Abl32767 Human imm
c	30	40	3.5	6113	6	ABL32802	Abl32802 Human imm
c	31	40	3.5	6127	6	ABL32593	Abl32593 Human imm
c	32	39.8	3.4	15165	4	ABL20844	Abl20844 Drosophil
	33	39.6	3.4	1992	8	ACA28307	Aca28307 Prokaryot
c	34	39.4	3.4	2324	2	AAX20302	Aax20302 Borrelia
c	35	39.4	3.4	4031	4	ABL27128	Abl27128 Drosophil
	36	39.4	3.4	6203	4	AAS45476	Aas45476 Chemicall
	37	39.4	3.4	6203	6	ABK28400	Abk28400 DNA trans
c	38	39.4	3.4	7175	2	AAV35027	Aav35027 Oryza sat
c	39	39.4	3.4	8923	4	AAS46339	Aas46339 Tumour su
c	40	39	3.4	1572	13	ADX30264	Adx30264 Plant ful
c	41	39	3.4	1689	2	AAQ66824	Aaq66824 AmEPV NPH
c	42	39	3.4	1689	2	AAZ10091	Aaz10091 Amsacta m
c	43	39	3.4	1689	6	ABT08025	Abt08025 Partial A
	44	39	3.4	1947	6	ABL56231	Abl56231 AmEPV NPH
c	45	39	3.4	1947	6	ABT08052	Abt08052 AmEPV NPH

ALIGNMENTS

RESULT 1
AAX06750
ID AAX06750 standard; DNA; 12441 BP.
XX
AC AAX06750;
XX
DT 26-APR-1999 (first entry)
XX
DE Salmonella enterica serovar C2 O antigen gene cluster.
XX
KW O antigen; diagnosis; wbaR gene; wzx gene; wbaL gene; wbaQ gene;
KW wxy gene; wbaW gene; wbaZ gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Salmonella enterica.
XX
FH Key Location/Qualifiers
FT CDS 138. .1019
FT /*tag= a
FT /note= "abe gene; encodes AAW88324"
FT CDS 1019. .2359
FT /*tag= b
FT /product= "flippase"
FT /note= "wzx gene, this region is specifically claimed in
FT Claim 9; encodes AAW88325"
FT CDS 2362. .3314
FT /*tag= c
FT /product= "abequosyl_transferase"
FT /note= "wbaR gene, this region is specifically claimed in
FT Claim 9; encodes AAW88326"
FT CDS 3361. .3975
FT /*tag= d
FT /product= "acetyl_transferase"
FT /note= "wbaL gene, this region is specifically claimed in
FT Claim 9; encodes AAW88327"
FT CDS 3977. .5020
FT /*tag= e
FT /product= "rhamnosyl_transferase"
FT /note= "wbaQ gene, this region is specifically claimed in
FT Claim 9; encodes AAW88328"
FT CDS 5114. .6313
FT /*tag= f

```
FT                   /product= "O_antigen_polymerase"
FT                   /note= "wzy gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88329"
FT   CDS             6313. .7323
FT                   /*tag= g
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88330"
FT   CDS             7310. .8467
FT                   /*tag= h
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88331"
FT   CDS             8482. .9903
FT                   /*tag= i
FT                   /note= "manC gene; encodes AAW88332"
FT   CDS             9911. .11347
FT                   /*tag= j
FT                   /note= "manB gene; encodes AAW88333"
FT   CDS             11426. .12441
FT                   /*tag= k
FT                   /note= "wbaP gene; encodes AAW88334"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;    98WO-AU000315.
XX
PR   01-MAY-1997;    97AU-00006545.
PR   22-JUL-1997;    97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88324, AAW88325, AAW88326, AAW88327, AAW88328, AAW88329,
DR   AAW88330, AAW88331, AAW88332, AAW88333, AAW88334.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 9; Fig 9; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Salmonella enterica serotype C2 (strain M67) O
CC   antigen. 11 Open reading frames within the gene cluster were identified
CC   on the basis of homology to known sequences. The genes encode proteins
CC   (see AAW88323-34) involved the synthesis of sugars present in the
CC   polysaccharide antigen, and in the transport or processing of
CC   polysaccharide or oligosaccharide units. The use of nucleic acid
CC   molecules derived from assembly and transport genes, particularly wbd
CC   (transferase), wzx (flippase) and wzy (polymerase) genes, within O
CC   antigen gene clusters improves the specificity of methods for the
CC   detection and identification of O antigens, e.g. in testing food- or
CC   faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of S. enterica
XX
SQ   Sequence 12441 BP; 3887 A; 1876 C; 2518 G; 4160 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1158;  DB 2;  Length 12441;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-280;
  Matches 1158;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7310 ATGGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGC 7369

Qy         61 TCCGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7370 TCCGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTG 7429

Qy        121 GTTGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7430 GTTGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACA 7489

Qy        181 TTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7490 TTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCA 7549

Qy        241 CTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7550 CTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCC 7609

Qy        301 GTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCT 7669

Qy        361 ATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 ATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGA 7729

Qy        421 ATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 ATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACC 7789

Qy        481 GCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAA 7849

Qy        541 GTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 GTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTA 7909

Qy        601 AAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 AAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATT 7969

Qy        661 GATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGAT 8029

Qy        721 GGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTAT 8089

Qy        781 CAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCG 8149

Qy        841 GAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCC 8209

Qy        901 TTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGC 8269

Qy        961 ATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 ATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAA 8329

Qy       1021 AATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 AATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCA 8389

Qy       1081 AGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAA 8449

Qy       1141 CAGATTATTAAACGTTAA 1158
              ||||||||||||||||||
Db       8450 CAGATTATTAAACGTTAA 8467
```

RESULT 2
ACH95204
ID ACH95204 standard; DNA; 1167 BP.
XX
AC ACH95204;
XX
DT 29-JUL-2004 (first entry)
XX
DE Klebsiella pneumoniae polynucleotide seqid 999.
XX
KW Recombinant expression vector; transcription regulatory element;
KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.
XX
OS Klebsiella pneumoniae.
XX
PN US6610836-B1.
XX
PD 26-AUG-2003.
XX
PF 27-JAN-2000; 2000US-00489039.
XX
PR 29-JAN-1999; 99US-0117747P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton GL, Osborne M;
XX
DR WPI; 2003-895346/82.
DR P-PSDB; ABO61653.
XX
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS Disclosure; SEQ ID NO 999; 932pp; English.
XX
CC The invention describes a new isolated nucleic acid encoding a Klebsiella
CC pneumoniae polypeptide. Also described are: a recombinant expression
CC vector comprising the nucleic acid, operably linked to a transcription
CC regulatory element; and a cell comprising the recombinant expression
CC vector. The nucleic acid is useful for preparing a vaccine composition
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella
CC pneumoniae polypeptide of the invention
XX
SQ Sequence 1167 BP; 371 A; 181 C; 250 G; 365 T; 0 U; 0 Other;

Query Match 44.9%; Score 520; DB 11; Length 1167;
Best Local Similarity 65.7%; Pred. No. 3.1e-120;
Matches 757; Conservative 0; Mismatches 395; Indels 0; Gaps 0;

```
Qy          7 ATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTCCGAA 66
              | ||  || || || ||   | ||||   ||| |||||  | | |||||| || || |||
Db         16 AAAACGGACATTAACATTGGTGTCGTGGCTGATTGGTTTATCACTTATGCTGGTTCAGAA 75

Qy         67 CAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGTTGAT 126
               ||||     |||   |  |   | | |||||| ||    || | ||  | || || |||
Db         76 AAGGTTGTCGCTGAGTTTTTAAGTTTATTTCCTAAATGCGAGCTTTACGCTGTTGTGGAT 135

Qy        127 TTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATTTATT 186
              |||||  |  | ||  | | |  |||||||| |   |||      || || || || |||
Db        136 TTTCTCTCTCAAGAGAATAAATTACATTTTCAGAATAAAGTAATAACAACTACTTTCATT 195

Qy        187 CAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCT 246
              |||||  || ||||||||    | |   ||||||  ||| |||||| | ||||||||||
Db        196 CAAAACTTGCCTAAAGCACGTACAAAATACCAGTCATACTTACCATTTATGCCACTGGCA 255

Qy        247 ATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCA 306
              ||||||||  ||||||  | ||   |   | || | ||  ||  |||   |||| | || ||
Db        256 ATTGAACAGCTTGATGTCTCTAAGCATGATGTAATACTTCTTAGTAGCCATGCAGTCGCT 315

Qy        307 AAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTATTCGA 366
              || || ||  || | ||||| || || |||||||| || ||||||||||||||||||||
Db        316 AAGGGCGTGCTTACAGGACCTGACCAACTTCACATAAGTTATGTTCATTCTCCTATTCGT 375

Qy        367 TATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAA 426
              ||||| |||||||| || |||||||||  | |  ||| ||    ||| ||| ||| | |||
Db        376 TATGCATGGGATTTGCAACATCAGTACTTGAGAGAGGCTGGATTGAGTAAGGGAGTGAAA 435

Qy        427 GGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAAT 486
              | |   || ||    ||| |  | || |||||| | || ||||||| || || ||||||
Db        436 GCTCTCTTGGCTCGGTGGATATTACATAAAATAAGGATGTGGGATTCCCGGACTGCAAAT 495

Qy        487 GGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATAC 546
              || || || ||||| || ||||||||| | | ||| |||||| | || |||||||| ||
Db        496 GGTGTAGACCATTTCATTGCTAATTCTAATTTTATTGCGCGTCGTATAAAAAAAGTCTAT 555

Qy        547 AGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAAT 606
              |  |  | ||    |||||||||||| ||||| ||||| || || ||   | |  || |
Db        556 GGCAGGAATGCGGATGTTATATATCCTCCTGTTGATGTCGACAACTTCACATTTCAAGAA 615

Qy        607 GAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTT 666
              || ||| |||| |||||||| || ||||| ||| |||| || || |||||||||||| |
Db        616 GAGAAGGAAGAATATTATTTTACGGCATCTCGTTTGGTTCCGTATAAACGTATTGATTTG 675

Qy        667 ATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCG 726
              ||||| || || ||||||   ||||| | ||  ||| | || ||| | ||||| |||||
Db        676 ATTGTTGAGGCATTTAGTTGTATGCCAGGAAGAAAACTTGTTGTTGTCGGTGACGGACCC 735

Qy        727 GAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTATCAATCT 786
              || |||| ||| ||||| |  || || ||| | || ||| |   | || || || || || ||
Db        736 GAAATGAGAAAGATAAAAAATAAAGCAACAAAAAACATAGATATTCTAGGCTACCAGTCA 795

Qy        787 TTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAG 846
                |  | | |   |||||||||||  |  ||  |  || |||||||||| || |||||
Db        796 AATGAGGCTATGCGGGAGTATATGCAAAAAGCAAAAGCATTTGTTTTTGCGGCCGAAGAA 855

Qy        847 GACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGG 906
              || ||||| || |  || || || || || || || || || ||||||||||||||||||
Db        856 GATTTTGGCATTACCCCGGTTGAGGCCCAGGCGTGTGGAACACCTGTTATTGCCTTTGGT 915

Qy        907 AAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGCATTTTC 966
              || |||||  | || ||||| ||  | |||   || |||    | || ||||| |||||
Db        916 AAAGGTGGATCTTTGGAAACAGTAAGACCATATGGAGTAAGCAATCCTACTGGTATTTTA 975

Qy        967 TTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAAAATGCA 1026
              ||  |  | ||  ||||    ||| |   | | | ||||  |  |||||||
Db        976 TTTGAAAACCAAGATATACAATCTATAATTAACGGTGTTGATTTATTTGAACGTGTATGC 1035

Qy       1027 TCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCAAGATTT 1086
                  |||   ||   | |||||||||  |||| | |||||||||       ||||||
Db       1036 GATGAAATTAGTCCTGGAGATTGTAGAAAGCATGCTGTAAAATTTTCCATTGATAGATTT 1095

Qy       1087 GAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAACAGATT 1146
                  | ||| | ||  | | ||||   ||||| |  ||    |||  |   |   | |
Db       1096 CGTAATGAAATAAATGATTATGTTCTAGAAAAATATAACAAATTCTTAGATGCGAAAAAA 1155

Qy       1147 ATTAAACGTTAA 1158
              || ||  ||||
Db       1156 ATAAATTATTAA 1167
```

RESULT 3
ACH95410/c
ID ACH95410 standard; DNA; 324 BP.
XX
AC ACH95410;
XX
DT 29-JUL-2004 (first entry)
XX
DE Klebsiella pneumoniae polynucleotide seqid 1205.
XX
KW Recombinant expression vector; transcription regulatory element;
KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.
XX
OS Klebsiella pneumoniae.
XX

PN US6610836-B1.
XX
PD 26-AUG-2003.
XX
PF 27-JAN-2000; 2000US-00489039.
XX
PR 29-JAN-1999; 99US-0117747P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton GL, Osborne M;
XX
DR WPI; 2003-895346/82.
DR P-PSDB; ABO61859.
XX
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS Disclosure; SEQ ID NO 1205; 932pp; English.
XX
CC The invention describes a new isolated nucleic acid encoding a Klebsiella
CC pneumoniae polypeptide. Also described are: a recombinant expression
CC vector comprising the nucleic acid, operably linked to a transcription
CC regulatory element; and a cell comprising the recombinant expression
CC vector. The nucleic acid is useful for preparing a vaccine composition
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella
CC pneumoniae polypeptide of the invention
XX
SQ Sequence 324 BP; 99 A; 70 C; 46 G; 109 T; 0 U; 0 Other;

Query Match 15.7%; Score 181.4; DB 11; Length 324;
Best Local Similarity 73.0%; Pred. No. 2.1e-35;
Matches 233; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```
Qy        449 TTCACAAAATACGAATTTGGGATTCTCGAACCGCAAATGGGGTTGATCATTTTATAGCTA 508
              | || |||||| | || |||||||| || || |||||||| || || ||||| || ||||
Db        323 TACATAAAATAAGGATGTGGGATTCCCGGACTGCAAATGGTGTAGACCATTTCATTGCTA 264

Qy        509 ATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGGCTTCAGTTATAT 568
              ||||| | | ||| |||||| | || |||||||| ||  |  |  | ||    |||||||
Db        263 ATTCTAATTTTATTGCGCGTCGTATAAAAAAAGTCTATGGCAGGAATGCGGATGTTATAT 204

Qy        569 ATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAGACTATTATTTCA 628
              |||| ||||| ||||| || || ||    | |  || | || ||| |||| |||||||| |
Db        203 ATCCTCCTGTTGATGTCGACAACTTCACATTTCAAGAAGAGAAGGAAGAATATTATTTTA 144

Qy        629 CAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAAGCCTTTAGTAAAA 688
              | ||||| ||| |||| || || ||||||||||| | ||||| || || ||||||    |
Db        143 CGGCATCTCGTTTGGTTCCGTATAAACGTATTGATTTGATTGTTGAGGCATTTAGTTGTA 84

Qy        689 TGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAAAAAATAAAGAGCA 748
              |||| | ||  ||| | || ||| | ||||| ||||| || |||| ||| ||||| |  |
Db         83 TGCCAGGAAGAAAACTTGTTGTTGTCGGTGACGGACCCGAAATGAGAAAGATAAAAAATA 24

Qy        749 AGGCTACAGACAATATAAA 767
              | || ||| | || ||| |
Db         23 AAGCAACAAAAAACATAGA 5
```

RESULT 4
ACH95331/c
ID ACH95331 standard; DNA; 222 BP.
XX
AC ACH95331;
XX
DT 29-JUL-2004 (first entry)
XX
DE Klebsiella pneumoniae polynucleotide seqid 1126.
XX
KW Recombinant expression vector; transcription regulatory element;
KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.
XX
OS Klebsiella pneumoniae.

XX
PN US6610836-B1.
XX
PD 26-AUG-2003.
XX
PF 27-JAN-2000; 2000US-00489039.
XX
PR 29-JAN-1999; 99US-0117747P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton GL, Osborne M;
XX
DR WPI; 2003-895346/82.
DR P-PSDB; ABO61780.
XX
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS Disclosure; SEQ ID NO 1126; 932pp; English.
XX
CC The invention describes a new isolated nucleic acid encoding a Klebsiella
CC pneumoniae polypeptide. Also described are: a recombinant expression
CC vector comprising the nucleic acid, operably linked to a transcription
CC regulatory element; and a cell comprising the recombinant expression
CC vector. The nucleic acid is useful for preparing a vaccine composition
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella
CC pneumoniae polypeptide of the invention
XX
SQ Sequence 222 BP; 68 A; 53 C; 44 G; 57 T; 0 U; 0 Other;

Query Match 10.4%; Score 120.2; DB 11; Length 222;
Best Local Similarity 71.5%; Pred. No. 4.4e-20;
Matches 158; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        226 TTACCACTAATGCCACTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATT 285
              |||||| | |||||||||||| |||||||| |||||| | ||  |   | || | ||  ||
Db        222 TTACCATTTATGCCACTGGCAATTGAACAGCTTGATGTCTCTAAGCATGATGTAATACTT 163

Qy        286 AGTAGCGCCCATTCCGTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGC 345
                |||   |||| | || || || || ||  || | ||||| || || ||||||||  ||
Db        162 CTTAGTAGCCATGCAGTCGCTAAGGGCGTGCTTACAGGACCTGACCAACTTCACATAAGT 103

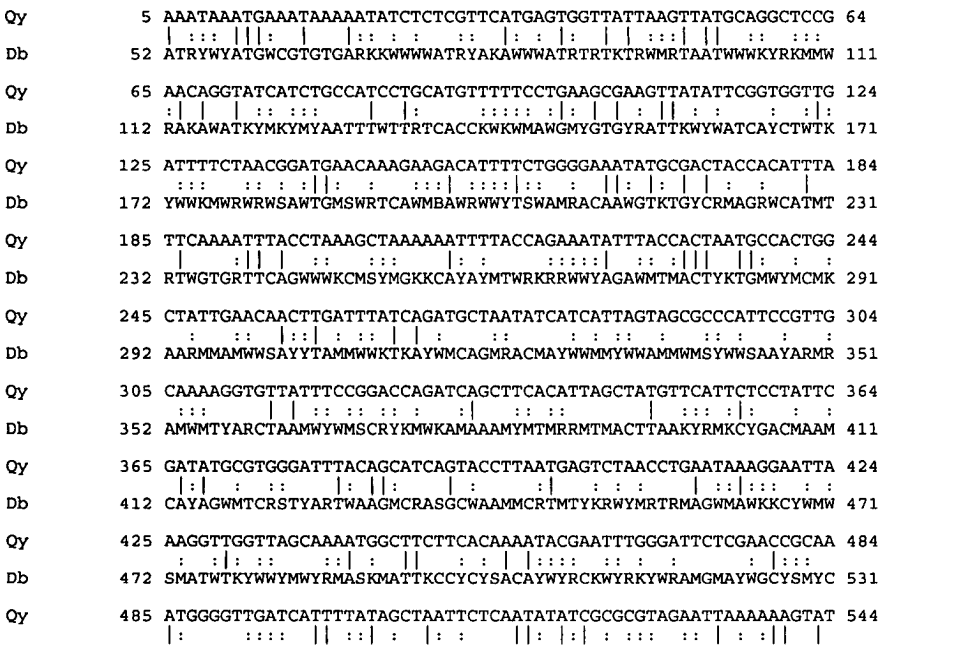
Qy        346 TATGTTCATTCTCCTATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCT 405
              |||||||||||||||||| ||||| |||||||| || ||||||||  | |  ||| ||
Db        102 TATGTTCATTCTCCTATTCGTTATGCATGGGATTTGCAACATCAGTACTTGAGAGAGGCT 43

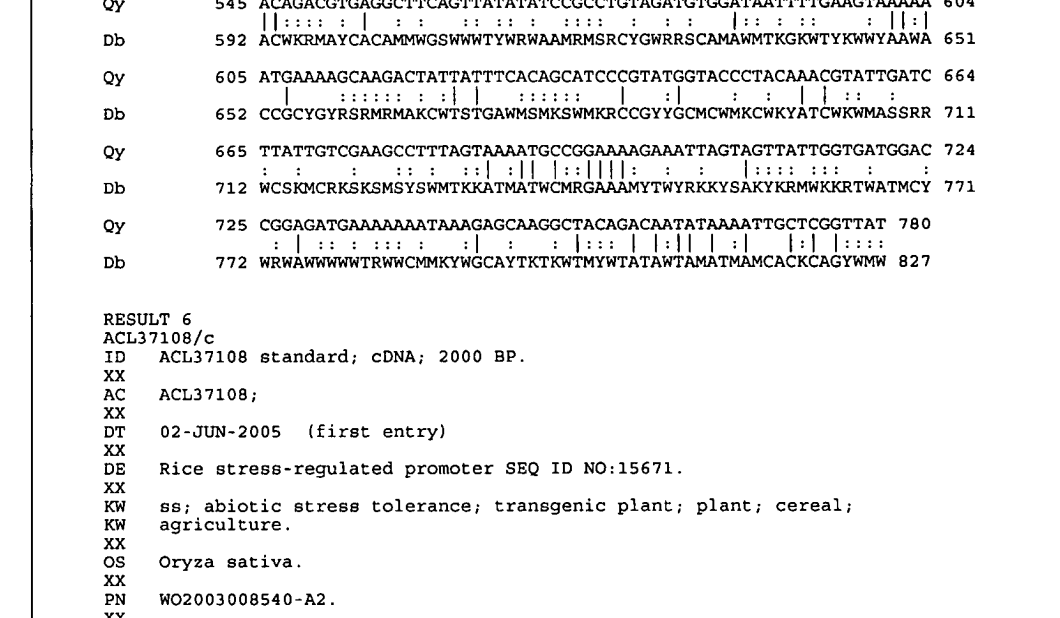
Qy        406 AACCTGAATAAAGGAATTAAAGGTTGGTTAGCAAAATGGCT 446
                ||| ||| ||| | |||| |   || ||    ||| |
Db         42 GGATTGAGTAAGGGAGTGAAAGCTCTCTTGGCTCGGTGGAT 2
```

RESULT 5
ACL35887
ID ACL35887 standard; cDNA; 2000 BP.
XX
AC ACL35887;
XX
DT 02-JUN-2005 (first entry)
XX
DE Rice stress-regulated promoter SEQ ID NO:14450.
XX
KW ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW agriculture.
XX
OS Oryza sativa.
XX
PN WO2003008540-A2.
XX
PD 30-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US019668.
XX

PR 22-JUN-2001; 2001US-0300112P.
PR 24-AUG-2001; 2001US-0314662P.
PR 26-SEP-2001; 2001US-0325277P.
PR 21-NOV-2001; 2001US-0332132P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
PI Moughamer T, Provart N, Ricke D, Zhu T;
XX
DR WPI; 2003-248011/24.
XX
PT New stress-responsive nucleic acid, useful for altering the
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT stress, salt stress or osmotic stress.
XX
PS Claim 48; SEQ ID NO 14450; 89pp; English.
XX
CC The invention relates to novel abiotic stress responsive polynucleotides
CC and polypeptides. Also disclosed are vectors, expression cassettes, host
CC cells, and plants containing such polynucleotides. Also disclosed are
CC methods for using the polynucleotides and polypeptides to alter the
CC responsiveness of a plant to abiotic stress. The invention is useful in
CC agriculture. The nucleic acid is useful for determining whether a test
CC plant has been exposed to an abiotic stress condition. It is also useful
CC for selecting an agent that alters abiotic stress regulated
CC polynucleotide expression in a plant cell, and to identify a homolog or
CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC molecule and the polypeptide encoded by it are useful in altering the
CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC stress, osmotic stress or any of their combinations. The present sequence
CC is used in the exemplification of the invention
XX
SQ Sequence 2000 BP; 410 A; 329 C; 253 G; 416 T; 0 U; 592 Other;

Query Match 4.7%; Score 54; DB 11; Length 2000;
Best Local Similarity 15.2%; Pred. No. 0.0038;
Matches 118; Conservative 276; Mismatches 382; Indels 0; Gaps 0;





RESULT 6
ACL37108/c
ID ACL37108 standard; cDNA; 2000 BP.
XX
AC ACL37108;
XX
DT 02-JUN-2005 (first entry)
XX
DE Rice stress-regulated promoter SEQ ID NO:15671.
XX
KW ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW agriculture.
XX
OS Oryza sativa.
XX
PN WO2003008540-A2.
XX
PD 30-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US019668.
XX
PR 22-JUN-2001; 2001US-0300112P.
PR 24-AUG-2001; 2001US-0314662P.
PR 26-SEP-2001; 2001US-0325277P.
PR 21-NOV-2001; 2001US-0332132P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
PI Moughamer T, Provart N, Ricke D, Zhu T;
XX
DR WPI; 2003-248011/24.
XX
PT New stress-responsive nucleic acid, useful for altering the
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT stress, salt stress or osmotic stress.
XX
PS Claim 48; SEQ ID NO 15671; 89pp; English.
XX
CC The invention relates to novel abiotic stress responsive polynucleotides
CC and polypeptides. Also disclosed are vectors, expression cassettes, host
CC cells, and plants containing such polynucleotides. Also disclosed are
CC methods for using the polynucleotides and polypeptides to alter the
CC responsiveness of a plant to abiotic stress. The invention is useful in
CC agriculture. The nucleic acid is useful for determining whether a test
CC plant has been exposed to an abiotic stress condition. It is also useful
CC for selecting an agent that alters abiotic stress regulated
CC polynucleotide expression in a plant cell, and to identify a homolog or
CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC molecule and the polypeptide encoded by it are useful in altering the
CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC stress, osmotic stress or any of their combinations. The present sequence
CC is used in the exemplification of the invention
XX

```
SQ   Sequence 2000 BP; 510 A; 350 C; 268 G; 518 T; 0 U; 354 Other;

  Query Match             4.3%;  Score 50.2;  DB 11;  Length 2000;
  Best Local Similarity   14.1%;  Pred. No. 0.034;
  Matches   70;  Conservative  217;  Mismatches  201;  Indels   10;  Gaps    2;

Qy         90 TGTTTTTCCTGAAGCGAAGTTATATTCGGTGGTTGATTTTCTAACGGATGAACAAAGAAG 149
              ||  :  ||||||: | |::||::  : :::||::  :::    :| : : :|: |  ::
Db        494 TGNKAGTCCTGARACCAWRTKWCKWMMRKTGKKRWWYKGMKRMARKAWGKARWAYCKRTT 435

Qy        150 ACATTTTCTGGGGAAATATGCGACTACCACATTTATTCAAAATTTACCTAAAGCTAAAAA 209
              |:|| ||  |:: :||:::   :  : :: ::      : :::::::: ::: | :|:
Db        434 AYATMTTTTRRTRAAWWMYTGMRKRRRSMTRWRMTSRWTTWSWYKWYYYWTRRYRASRAR 375

Qy        210 ATTTTACCAGAAATATTTACCACTAATGCCACTGGCTATTGAACAACTTGATTTATCAGA 269
               :       | | : | |:  ::  :    | :  :  :::: :|:: :   | ::: ::
Db        374 TK-------GKATKYTMTWKWMMGCWSYTKARWWMMAYKKSWKMAWYAWTKGTRRYYYRM 322

Qy        270 TGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCAAAAGGTGTTATTTCCGGACCAGA 329
              ::   :|:::||   |: |      :|:|   :      |   | : ::|:: ||:  :
Db        321 KRAAYWTRWMATRSATWCTGWTYWTSAKTTKWKRWAGYAKATTMKMRKTYYAGGRWWWYK 262

Qy        330 TCAGC---TTCACATTAGCTATGTTCATTCTCCTATTCGATATGCGTGGGATTTACAGCA 386
              : |::   ::|::  ::  | :::::: | :  :  ::: ::| : :  : ::| : ::
Db        261 KWASSRTWKWCWYKMYMWATKKSWWSWATTWAWKSMWYKKKWTYYTYMTSTYWTTWMWMR 202

Qy        387 TCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAAGGTTGGTTAGCAAAATGGCT 446
               ||::    :|::|   | :: | ::: ::: ::::: : : :: :  ::   ::  ::
Db        201 CRAGWWWKGGMAKRAAARTKWMRTYRMKKMWRYRWYKWSWARGWRAWRKRYYTYWWYMMY 142

Qy        447 TCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAATGGGGTTGATCATTTTATAGC 506
               | :::: | :   :::|||: :  :::  ::: :  |: :  : : ::      |:: |
Db        141 RCAWMWYSAYWRSWRRWTTTKYSKMYYWKMWMMRRTYARAKAYSRWWMWKKRCRTRWKGW 82

Qy        507 TAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGGCTTCAGTTAT 566
              |:::| :::: :: :  :::::   :::::  | : ::: : |:: ::: : :   :: :
Db         81 TRWKTWKMMWRWYKKRTYRMKKSAKWWYMRSGAYRRRWMRRMAYSRRRSWMRKTSYYYTK 22

Qy        567 ATATCCGCCTGTAGATGT 584
               ::: : ::  ::: : |
Db         21 KWWKKMWSSGYYWRYYTT 4


RESULT 7
ACL35363/c
ID   ACL35363 standard; cDNA; 2000 BP.
XX
AC   ACL35363;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Rice stress-regulated promoter SEQ ID NO:13926.
XX
KW   ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW   agriculture.
XX
OS   Oryza sativa.
XX
PN   WO2003008540-A2.
XX
PD   30-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US019668.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   24-AUG-2001; 2001US-0314662P.
PR   26-SEP-2001; 2001US-0325277P.
PR   21-NOV-2001; 2001US-0332132P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Kreps J,  Briggs SP,  Cooper B,  Glazebrook J,  Goff SA,  Katagiri F;
PI   Moughamer T,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-248011/24.
XX
PT   New stress-responsive nucleic acid, useful for altering the
PT   responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT   stress, salt stress or osmotic stress.
XX
PS   Claim 48; SEQ ID NO 13926; 89pp; English.
XX
CC   The invention relates to novel abiotic stress responsive polynucleotides
CC   and polypeptides. Also disclosed are vectors, expression cassettes, host
CC   cells, and plants containing such polynucleotides. Also disclosed are
CC   methods for using the polynucleotides and polypeptides to alter the
CC   responsiveness of a plant to abiotic stress. The invention is useful in
CC   agriculture. The nucleic acid is useful for determining whether a test
CC   plant has been exposed to an abiotic stress condition. It is also useful
CC   for selecting an agent that alters abiotic stress regulated
CC   polynucleotide expression in a plant cell, and to identify a homolog or
CC   ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC   molecule and the polypeptide encoded by it are useful in altering the
CC   responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC   stress, osmotic stress or any of their combinations. The present sequence
CC   is used in the exemplification of the invention
XX
SQ   Sequence 2000 BP; 540 A; 344 C; 416 G; 428 T; 0 U; 272 Other;

  Query Match             4.1%;  Score 47;  DB 11;  Length 2000;
  Best Local Similarity   12.9%;  Pred. No. 0.22;
  Matches   47;  Conservative  158;  Mismatches  158;  Indels    0;  Gaps    0;

Qy        447 TCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAATGGGGTTGATCATTTTATAGC 506
              || | |   ||  ||  |  :  ::| |  ::    ::: : :: :    |: ::::
Db        378 TCGTAAAGGAAGTCGTTTAGSWCMWTKTWWMWKTMRMWMMKTKKAWAKAWAWAYYMYSMR 319

Qy        507 TAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGGCTTCAGTTAT 566
              ::: :::::| :: : : :  |: ::: :|:::  :|:::: |:::  :: ::  :  :
Db        318 YRWMYYYSWAAWKGWWRWSKTTWARRWSWAWRWKTWAKWYRAAYSKWKKKWWYWTRAAWM 259

Qy        567 ATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAGACTATTATTT 626
              ||:    ::| :: :: |: ::| :|:  :: : ||:||: | : : | :  ::| : :
Db        258 ATRMWTWMMTWWWARKWTRYWWAKWTYAMWWTKYAAWAAWTATWTKWATRGAWWTRWSWA 199

Qy        627 CACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAAGCCTTTAGTAA 686
               ::|   ::  :::| : |  ::::: :  | :   :   ::  ::     : :    :
Db        198 GWMAAAKWYWTKYWTWRRAGWMYRYRTMKMTSWCYTWWMMRWMWWYWTTTWYRWRKAMRK 139

Qy        687 AATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAAAAAATAAAGAG 746
              |: :    ::::: |: :  | : :::  :  :    :| :: : ::     :::: :::
Db        138 ARGRAGTMWWWRKYAWSWTWTKKRKWYMWKRTMRYWKYCCKRAWMSWCYTKSWWWWYKMS 79

Qy        747 CAAGGCTACAGACAATATAAAATTGCTCGGTTATCAATCTTTTCCTGTTTTAAAAGAGTA 806
               : :::  :|| :  :|: ::: |: : ::|::|::   : | |: : |   :::: : :
Db         78 TWSKRMSGSAGYMTKYAKYWMMMTRWKTRKTKRTSWKAWWMTCCMRKRTMCKMMWKKRMW 19

Qy        807 TAT 809
              :::
Db         18 YMK 16


RESULT 8
ACL37108
ID   ACL37108 standard; cDNA; 2000 BP.
XX
AC   ACL37108;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Rice stress-regulated promoter SEQ ID NO:15671.
XX
KW   ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW   agriculture.
XX
OS   Oryza sativa.
```

XX
PN WO2003008540-A2.
XX
PD 30-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US019668.
XX
PR 22-JUN-2001; 2001US-0300112P.
PR 24-AUG-2001; 2001US-0314662P.
PR 26-SEP-2001; 2001US-0325277P.
PR 21-NOV-2001; 2001US-0332132P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
PI Moughamer T, Provart N, Ricke D, Zhu T;
XX
DR WPI; 2003-248011/24.
XX
PT New stress-responsive nucleic acid, useful for altering the
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT stress, salt stress or osmotic stress.
XX
PS Claim 48; SEQ ID NO 15671; 89pp; English.
XX
CC The invention relates to novel abiotic stress responsive polynucleotides
CC and polypeptides. Also disclosed are vectors, expression cassettes, host
CC cells, and plants containing such polynucleotides. Also disclosed are
CC methods for using the polynucleotides and polypeptides to alter the
CC responsiveness of a plant to abiotic stress. The invention is useful in
CC agriculture. The nucleic acid is useful for determining whether a test
CC plant has been exposed to an abiotic stress condition. It is also useful
CC for selecting an agent that alters abiotic stress regulated
CC polynucleotide expression in a plant cell, and to identify a homolog or
CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC molecule and the polypeptide encoded by it are useful in altering the
CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC stress, osmotic stress or any of their combinations. The present sequence
CC is used in the exemplification of the invention
XX
SQ Sequence 2000 BP; 510 A; 350 C; 268 G; 518 T; 0 U; 354 Other;

Query Match 3.9%; Score 45; DB 11; Length 2000;
Best Local Similarity 12.5%; Pred. No. 0.69;
Matches 52; Conservative 188; Mismatches 168; Indels 9; Gaps 1;

```
Qy        676 GCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAA 735
Db         59 SMMKYRAYMMRWYWKKMWAMWYAWCMWYAYGYMMWKWWYSRTMTYTRAYYKKWKMWRRKM 118

Qy        736 AAAATAAAGAGCAAGGCTACAGACAATATATAAAATTGCTCGGTTATCAATCTTTTCCTGTT 795
Db        119 SRMAAAWYYWSYWRTSRWKWTGYRKKRWWRARRYMYWTYWCYTWSWMRWYRYWKMMKYRA 178

Qy        796 TTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGA 855
Db        179 YKWMAYTTTYMT---------KCCMWWWCTYGYKWKWAAWRASAKRARRAWMMMRWKSMW 229

Qy        856 ATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGG 915
Db        230 TWAATWSWWSMMATWKRKMRWGWMWAYSSTWMMRWWWYCCTRRAMYKMKAATMTRCTWYM 289

Qy        916 GCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAA 975
Db        290 WMAAMTSAWRAWCAGWATSYATKWYAWRTTYMKYRRRYYACMAWTRWTKMWSMMRTKKWW 349

Qy        976 CAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTT 1035
Db        350 YTMARSWGCKKWMWAKARMATMCMAYTYSTYRYYAWRRRWMRWSWAAWYSAKYWYAKSYY 409

Qy       1036 ACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCAAGATTTGAACAA 1092
Db        410 YMYKCARKWWTTYAYYAAAAKATRTAAYMGRTWYTMCWTMYTKYMKCMRWWYMMCAM 466
```

RESULT 9
ACL35887/c
ID ACL35887 standard; cDNA; 2000 BP.
XX
AC ACL35887;
XX
DT 02-JUN-2005 (first entry)
XX
DE Rice stress-regulated promoter SEQ ID NO:14450.
XX
KW ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW agriculture.
XX
OS Oryza sativa.
XX
PN WO2003008540-A2.
XX
PD 30-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US019668.
XX
PR 22-JUN-2001; 2001US-0300112P.
PR 24-AUG-2001; 2001US-0314662P.
PR 26-SEP-2001; 2001US-0325277P.
PR 21-NOV-2001; 2001US-0332132P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
PI Moughamer T, Provart N, Ricke D, Zhu T;
XX
DR WPI; 2003-248011/24.
XX
PT New stress-responsive nucleic acid, useful for altering the
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT stress, salt stress or osmotic stress.
XX
PS Claim 48; SEQ ID NO 14450; 89pp; English.
XX
CC The invention relates to novel abiotic stress responsive polynucleotides
CC and polypeptides. Also disclosed are vectors, expression cassettes, host
CC cells, and plants containing such polynucleotides. Also disclosed are
CC methods for using the polynucleotides and polypeptides to alter the
CC responsiveness of a plant to abiotic stress. The invention is useful in
CC agriculture. The nucleic acid is useful for determining whether a test
CC plant has been exposed to an abiotic stress condition. It is also useful
CC for selecting an agent that alters abiotic stress regulated
CC polynucleotide expression in a plant cell, and to identify a homolog or
CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC molecule and the polypeptide encoded by it are useful in altering the
CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC stress, osmotic stress or any of their combinations. The present sequence
CC is used in the exemplification of the invention
XX
SQ Sequence 2000 BP; 410 A; 329 C; 253 G; 416 T; 0 U; 592 Other;

Query Match 3.8%; Score 44; DB 11; Length 2000;
Best Local Similarity 14.4%; Pred. No. 1.2;
Matches 75; Conservative 198; Mismatches 249; Indels 0; Gaps 0;

```
Qy        293 CCCATTCCGTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTC 352
Db        608 SCWKKTGTGRTKYMWGTSTYGTTKMYSAYWKRMWSKKRWGRTKYATGWKYWGYTGWYCAY 549

Qy        353 ATTCTCCTATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGA 412
Db        548 WSAAMCWKMRKGTKTRTGRKSRGCWRTKCKTYWRMYRWMGYRWRTGTSRGRGGMAATKMS 489

Qy        413 ATAAAGGAATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATT 472
Db        488 TKYRWKRWWRMAWATKSWKWRGMMWTKWCTKYAYKRWYMRAKAYGKKTTWGCSTYGKCTT 429
```

```
Qy        473 CTCGAACCGCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAA 532
Db        428 WAYTRASYGAKWCTRTGKTTKGTCRGMKYRMTTAAGTKAKYYKAKRKTTTKTMWKMRYGS 369

Qy        533 TTAAAAAAGTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATT 592
Db        368 KWRWKTTAGYTRAKWKTYKYTRTTSWWRSKWKKTWWRKKWWRTKGTYKCTGKWRTMAMWW 309

Qy        593 TTGAAGTAAAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACA 652
Db        308 KKTARRTSWWKTKKYTTMKGKRWKCAMRAGTKAKWTCTRWWYYMYWAKRTRTGMMCKRSK 249

Qy        653 AACGTATTGATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTA 712
Db        248 GMWWWCTGAAYCACWAYAKATGWYCTKYGRCAMACWTTGTYKTWSARWWYWTVKWTGAYW 189

Qy        713 TTGGTGATGGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGC 772
Db        188 SKCAWTSWYWYWKMWWRMAWAGRTGATWRWMAATYRCACRKCWTKWMWMGGTGAYAAWAA 129

Qy        773 TCGGTTATCAATCTTTTCCTGTTTTAAAAGAGTATATGCAGA 814
Db        128 ATTRKRMKRMATWTMTYWKKMYRMWWWATTAYKWYAMAYAYA 87
```

```
RESULT 10
AAH24065
ID   AAH24065 standard; DNA; 4590 BP.
XX
AC   AAH24065;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Yeast AOD9604-associated DNA sequence, SEQ ID NO:1.
XX
KW   Human growth hormone analogue peptide; hGH; AOD9604; lipid metabolism;
KW   modulation; lipolysis stimulation; hormone-sensitive lipase stimulation;
KW   lipogenesis inhibition; acetyl CoA carboxylase inhibition; obesity;
KW   functional food; transgenic yeast; fat/lean ratio; food use; ds.
XX
OS   Saccharomyces cerevisiae.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    10
FT                   /*tag=  a
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3617
FT                   /*tag=  b
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3649
FT                   /*tag=  c
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3679
FT                   /*tag=  d
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3819
FT                   /*tag=  e
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3862
FT                   /*tag=  f
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3864
FT                   /*tag=  g
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3888
FT                   /*tag=  h
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3890
FT                   /*tag=  i
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3912
FT                   /*tag=  j
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3914
FT                   /*tag=  k
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3938
FT                   /*tag=  l
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3939
FT                   /*tag=  m
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3941
FT                   /*tag=  o
FT                   /note= "Represented as * in the specification"
FT   misc_feature    3943
FT                   /*tag=  p
FT                   /note= "Represented as * in the specification"
FT   misc_feature    4361
FT                   /*tag=  q
FT                   /note= "Represented as * in the specification"
XX
PN   WO200133977-A1.
XX
PD   17-MAY-2001.
XX
PF   06-NOV-2000; 2000WO-AU001362.
XX
PR   05-NOV-1999;   99AU-00003875.
XX
PA   (META-) METABOLIC PHARM LTD.
XX
PI   Belyea CI,  Ng FM,  Vaughan P;
XX
DR   WPI; 2001-328876/34.
XX
PT   New organisms containing nucleic acid encoding a growth hormone fragment
PT   which modulates lipid metabolism are useful to produce dietary aids for
PT   obesity and in the meat production industry.
XX
PS   Disclosure; Page 48-50; 54pp; English.
XX
CC   The invention relates to novel transgenic organisms useful in the
CC   production of functional food and drink products for the treatment or
CC   prevention of obesity via the regulation of lipid metabolism. The
CC   organisms comprise a polynucleotide encoding a growth hormone fragment
CC   capable of stimulating the activity of hormone-sensitive lipase (the key
CC   enzyme in lipolysis) and inhibiting acetyl CoA carboxylase (the key
CC   enzyme in lipogenesis). The growth hormone fragment preferably contains
CC   at least the disulphide-bonded loop of a mammalian growth hormone (but is
CC   not the full-length growth hormone) and is optionally linked to an
CC   epitope tag or heterologous fusion protein partner. The transgenic
CC   organism may be a microorganism used to produce a fermented product
CC   (e.g., yeast), or an edible plant or animal or cell thereof. Food or
CC   drink made using methods of the invention are used to modify fat/lean
CC   ratio, lipid metabolism or food use in a mammal. In particular, the food
CC   or drink products may be used to treat or prevent obesity, particularly
CC   in humans, and may also be used to improve the fat/lean ration of
CC   livestock raised for meat production. In the exemplification of the
CC   invention, the human growth hormone (hGH) fragment analogue AOD9604 was
CC   expressed in yeast, optionally fused to the FLAG epitope (AAB73625). The
CC   present sequence is described as a DNA sequence from yeast in the
CC   sequence listing, but is not further referred to in the specification
XX
SQ   Sequence 4590 BP; 661 A; 384 C; 127 G; 522 T; 0 U; 2896 Other;
```

```
  Query Match             3.7%;  Score 42.6;  DB 5;  Length 4590;
  Best Local Similarity   12.2%;  Pred. No. 3.6;
  Matches   53;  Conservative  195;  Mismatches  188;  Indels    0;  Gaps    0;

Qy        358 CCTATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAA 417
Db       3980 CHKHHDHDHKGHMCMHCCCMHKMDHAAAANRMACMHHHHADNGCKHADATARAGRAHNTH 4039

Qy        418 GGAATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGA 477
```

```
Db       4040 ADRRTRRAGNMBRNNMBRSCDMMCKMMMMKWYKNDBDVWRYYTGBBDRKNWNDYSMBSHM 4099

Qy        478 ACCGCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAA 537

Db       4100 WNDYSKDCWNDYSCTMATRCVRYSAVASDWNDYSCTMATRCVRYSAVASDWNDYSCTMAT 4159

Qy        538 AAAGTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAA 597

Db       4160 RCVRYSAVASDWNDYSCTMATRCVRYSAVASDWNDYSCTMATRCVRYSAVASDWNDYSCT 4219

Qy        598 GTAAAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGT 657

Db       4220 MATRCVRYSAVASDRKSKWWSKKWSSMGTMSNWRMANSYMBARACRRNWHHBBSDHMBSS 4279

Qy        658 ATTGATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGT 717

Db       4280 TDWRDDCSTMATSSCDTWNDYSWNDYSHHSSWNDYSNDNDSCWNDYSDMCRSTWRDRRSD 4339

Qy        718 GATGGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGT 777

Db       4340 GRTHHACKCHTTDGDANBDDDRTNTRYRSHARRSRBTABRACATNSADMTADMNSTDWRD 4399

Qy        778 TATCAATCTTTTCCTG 793

Db       4400 DSTWRDDCMNTCMNTS 4415


RESULT 11
AAX20248_03/c
Continuation (4 of 10) of AAX20248 from base 300001 (Borrelia burgdorferi polynucleotide
WP Sequence split into 10 fragments  LOCUS AAX20248 Accession Aax20248
WP     Fragment Name      Begin     End
WP     AAX20248_00            1    110000
WP     AAX20248_01       100001    210000
WP     AAX20248_02       200001    310000
WP     AAX20248_03       300001    410000
WP     AAX20248_04       400001    510000
WP     AAX20248_05       500001    610000
WP     AAX20248_06       600001    710000
WP     AAX20248_07       700001    810000
WP     AAX20248_08       800001    910000
WP     AAX20248_09       900001    910715

  Query Match             3.7%;  Score 42.6;  DB 2;  Length 110000;
  Best Local Similarity   51.5%;  Pred. No. 10;
  Matches  124;  Conservative    0;  Mismatches  114;  Indels    3;  Gaps    1;

Qy        590 ATTTTGAAGTAAAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCT 649

Db      89912 ATTTAAAAATTAGAGGCGAAGAGCTTAAATATTCTTTAGAAAACTTTATTGCCTCATATA 89853

Qy        650 ACAAACGTATTGATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAG 709

Db      89852 ATGATAAAATTCAAAATTTAGAATATGATTTGTCTAAAAATTTGGAAAACAAAACAGAAC 89793

Qy        710 TTATT---GGTGATGGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAA 766

Db      89792 TTATTCAAAGTTTCCGTTTAGATATTGAACAAAAAATGAAAGATGATAAAGAGAATTTTT 89733

Qy        767 AATTGCTCGGTTATCAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGT 826

Db      89732 ATTTAGATTTTACTAAAGAATTTTCTTCTAAAAAAAAAGATATGCAGAGCGAGATAGCTT 89673

Qy        827 T 827

Db      89672 T 89672


RESULT 12
ADG47988
ID   ADG47988 standard; DNA; 1756 BP.
XX
AC   ADG47988;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Human UDP-glucoronosyl transferase (UGT) gene.
XX
KW   Human; disease-associated gene; therapy; cancer; gene;
KW   UDP-glucoronosyl transferase; UGT; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             93. .1676
FT                   /*tag=  a
FT                   /product= "Human UDP-glucoronosyl transferase (UGT)"
XX
PN   US2003101002-A1.
XX
PD   29-MAY-2003.
XX
PF   04-SEP-2002; 2002US-00235994.
XX
PR   01-NOV-2000; 2000US-0245081P.
PR   01-NOV-2001; 2001US-00003608.
XX
PA   (BART/) BARTHA G T.
PA   (WALK/) WALKER M.
XX
PI   Bartha GT,  Walker M;
XX
DR   WPI; 2004-008989/01.
DR   P-PSDB; ADG47989.
XX
PT   Analyzing gene expression useful for identifying biomolecules by
PT   analyzing dual channel DNA microarray images to determine expression
PT   patterns of one or more disease-specific genes.
XX
PS   Example; SEQ ID NO 27; 47pp; English.
XX
CC   The invention relates to diseases-associated genes and polypeptides
CC   encoded by these genes. The invention also relates to methods for
CC   analysing gene expression patterns and methods for diagnosing, treating
CC   or preventing disease. The invention is useful for analysing gene
CC   expression. It is also useful for identifying biomolecules such as
CC   polynucleotides or polypeptides useful in the diagnosis, prognosis,
CC   treatment, prevention and evaluation of therapies for diseases such as
CC   cancer. The present sequence is human UDP-glucoronosyl transferase (UGT)
CC   gene. This sequence is used to illustrate the method of the invention.
XX
SQ   Sequence 1756 BP; 538 A; 353 C; 363 G; 502 T; 0 U; 0 Other;

  Query Match             3.7%;  Score 42.4;  DB 12;  Length 1756;
  Best Local Similarity   49.1%;  Pred. No. 3;
  Matches  112;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        883 GGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGT 942

Db        138 GGAACCACTCTTGGTGGGAATGTTTTGATTTGGCCAATGGAAGGTAGTCATTGGCTAAAT 197

Qy        943 GTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCT 1002

Db        198 GTTAAGATAATTATAGATGAGCTCATTAAAAAGGAGCATAATGTGACTGTCCTAGTTGCC 257

Qy       1003 GTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCA 1062

Db        258 TCTGGTGCACTTTTCATCACACCAACCTCTAACCCATCTCTGACATTTGAAATATATAAG 317

Qy       1063 GAAAAATTTTCTCGATCAAGATTTGAACAAGAATTTAAGAACTTTGTT 1110

Db        318 GTGCCCTTTGGCAAAGAAAGAATAGAAGGAGTAATTAAGGACTTCGTT 365


RESULT 13
```

```
AAS91380
ID   AAS91380 standard; cDNA; 1766 BP.
XX
AC   AAS91380;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27184.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27193.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 27184; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1766 BP; 542 A; 359 C; 363 G; 502 T; 0 U; 0 Other;

  Query Match             3.7%;  Score 42.4;  DB 5;  Length 1766;
  Best Local Similarity   49.1%;  Pred. No. 3;
  Matches  112;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        883 GGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGT 942
              || ||| || ||  ||     ||     |  | || |  | |||     ||      ||| |
Db        109 GGAACCACTCTTGGTGGGAATGTTTTGATTTGGCCAATGGAAGGTAGTCATTGGCTAAAT 168

Qy        943 GTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCT 1002
              ||  ||  |   |  |      || | ||  |  || ||| ||   || | |  |  ||
Db        169 GTTAAGATAATTATAGATGAGCTCATTAAAAAGGAGCATAATGTGACTGTCCTAGTTGCC 228

Qy       1003 GTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCA 1062
              | ||| | ||   |  |   || | | |    ||||| | | | |  || | ||
Db        229 TCTGGTGCACTTTTCATCACACCAACCTCTAACCCATCTCTGACATTTGAAATATATAAG 288

Qy       1063 GAAAAATTTTCTCGATCAAGATTTGAACAAGAATTTAAGAACTTTGTT 1110
              |    |||     |  |||| | |||  || | ||||| |||| |||
Db        289 GTGCCCTTTGGCAAAGAAAGAATAGAAGGAGTAATTAAGGACTTCGTT 336


RESULT 14
AAS91382
ID   AAS91382 standard; cDNA; 2406 BP.
XX
AC   AAS91382;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27186.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27195.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 27186; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2406 BP; 654 A; 552 C; 561 G; 639 T; 0 U; 0 Other;

  Query Match             3.7%;  Score 42.4;  DB 5;  Length 2406;
  Best Local Similarity   49.1%;  Pred. No. 3.3;
```

```
  Matches  112;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        883 GGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGT 942
              || ||| || ||  ||    ||    |  | || |  | |||     ||     ||| |
Db       1135 GGAACCACTCTTGGTGGGAATGTTTTGATTTGGCCAATGGAAGGTAGTCATTGGCTAAAT 1194

Qy        943 GTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCT 1002
              ||  ||   |   |  |       || | ||  |  || ||| ||   || | |  |  ||
Db       1195 GTTAAGATAATTATAGATGAGCTCATTAAAAAGGAGCATAATGTGACTGTCCTAGTTGCC 1254

Qy       1003 GTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCA 1062
                | ||| | ||   |  |   || | | |    |||||| | | | |  || | ||
Db       1255 TCTGGTGCACTTTTCATCACACCAACCTCTAACCCATCTCTGACATTTGAAATATATAGG 1314

Qy       1063 GAAAAATTTTCTCGATCAAGATTTGAACAAGAATTTAAGAACTTTGTT 1110
              |     |||     | |||| | |||  || | ||||| |||| |||
Db       1315 GTGCCCTTTGGCAAAGAAAGAATAGAAGGAGTAATTAAGGACTTCGTT 1362


RESULT 15
AAD57521
ID   AAD57521 standard; cDNA; 2491 BP.
XX
AC   AAD57521;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human enzyme (ENZM) cDNA #33.
XX
KW   Human; enzyme; ENZM; immune disorder; infection; myocardial infarction;
KW   gene therapy; anaemia; acquired immune deficiency syndrome; infection;
KW   reproductive disorder; cardiovascular; eye; cell proliferation; cancer;
KW   AIDS; allergy; asthma; Addison's disease; diabetes; goitre; impotence;
KW   infertility; atherosclerosis; metabolic disorder; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             93. .1544
FT                   /*tag=  a
FT                   /product= "Human enzyme (ENZM)"
XX
PN   WO2003052075-A2.
XX
PD   26-JUN-2003.
XX
PF   12-DEC-2002; 2002WO-US040161.
XX
PR   14-DEC-2001; 2001US-0340357P.
PR   20-DEC-2001; 2001US-0342962P.
PR   21-DEC-2001; 2001US-0343558P.
PR   22-JAN-2002; 2002US-0351107P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Chawla NK,  Lee SY,  Ring HZ,  Lee EA,  Forsythe IJ,  Khare R;
PI   Tran UK,  Kable AE,  Richardson TW,  Emerling BM,  Lindquist EA;
PI   Baughn MR,  Hafalia AJA,  Jin P,  Swarnakar A,  Li JX,  Marquis JP;
PI   Lee S,  Gorvad AE,  Sprague WW,  Becha SD,  Elliott VS;
XX
DR   WPI; 2003-533016/50.
DR   P-PSDB; AAE38241.
XX
PT   New human enzyme (ENZM) polypeptide, useful for preparing a composition
PT   for treating a disease associated with decreased expression or
PT   overexpression of ENZM e.g. cancer.
XX
PS   Claim 5; Page 334-335; 264pp; English.
XX
CC   The invention relates to human enzyme (ENZM) polypeptides and their
CC   corresponding polynucleotides. ENZM sequences are useful for preparing a
CC   composition for diagnosing or treating a disease or condition associated
CC   with decreased expression or overexpression of functional ENZM. The
CC   disorders include immune disorders (anaemia, allergy or asthma),
CC   infectious disorders (viral, fungal, parasitic or protozoal infection),
CC   immune deficiencies (acquired immune deficiency syndrome; AIDS),
CC   metabolic disorder (Addison's disease, diabetes or goitre), reproductive
CC   disorders (infertility or impotence), cardiovascular disorders
CC   (atherosclerosis or myocardial infarction), eye disorders and cell
CC   proliferative disorders (cancer). ENZM DNA is useful in gene therapy. The
CC   present sequence is human ENZM cDNA
XX
SQ   Sequence 2491 BP; 792 A; 467 C; 453 G; 779 T; 0 U; 0 Other;

  Query Match             3.7%;  Score 42.4;  DB 9;  Length 2491;
  Best Local Similarity   49.1%;  Pred. No. 3.3;
  Matches  112;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        883 GGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGT 942
              || ||| || ||  ||    ||    |  | || |  | |||     ||     ||| |
Db        138 GGAACCACTCTTGGTGGGAATGTTTTGATTTGGCCAATGGAAGGTAGTCATTGGCTAAAT 197

Qy        943 GTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCT 1002
              ||  ||   |   |  |       || | ||  |  || ||| ||   || | |  |  ||
Db        198 GTTAAGATAATTATAGATGAGCTCATTAAAAAGGAGCATAATGTGACTGTCCTAGTTGCC 257

Qy       1003 GTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCA 1062
                | ||| | ||   |  |   || | | |    |||||| | | | |  || | ||
Db        258 TCTGGTGCACTTTTCATCACACCAACCTCTAACCCATCTCTGACATTTGAAATATATAAG 317

Qy       1063 GAAAAATTTTCTCGATCAAGATTTGAACAAGAATTTAAGAACTTTGTT 1110
              |     |||     | |||| | |||  || | ||||| |||| |||
Db        318 GTGCCCTTTGGCAAAGAAAGAATAGAAGGAGTAATTAAGGACTTCGTT 365


Search completed: May 11, 2006, 09:42:50
Job time : 530.645 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 09:43:19 ; Search time 3607.95 Seconds
(without alignments)
15016.694 Million cell updates/sec

Title: US-09-423-093B-3_COPY_7310_8467
Perfect score: 1158
Sequence: 1 atggaaataaatgaaataaa.....aacagattattaaacgttaa 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	58.6	5.1	1101	10	CNS0039G	AL063921 Drosophil
	2	49.4	4.3	987	10	CNS014PQ	AL104456 Drosophil
c	3	49.2	4.2	1101	10	CNS017KX	AL108171 Drosophil
	4	47.8	4.1	825	5	BU747088	BU747088 CH3#009_C
	5	47.6	4.1	872	8	CV966702	CV966702 PC011E12
	6	46.6	4.0	872	5	BU279377	BU279377 603864720
	7	46.6	4.0	924	10	CNS02Q6S	AL208909 Tetraodon
c	8	46.2	4.0	582	3	BJ407454	BJ407454 BJ407454
	9	46.2	4.0	1101	10	CNS0039G	AL063921 Drosophil
	10	46	4.0	498	5	BW542444	BW542444 BW542444
c	11	46	4.0	635	3	BJ348545	BJ348545 BJ348545
c	12	46	4.0	678	3	BJ356306	BJ356306 BJ356306
c	13	45.8	4.0	930	10	CZ974493	CZ974493 175383 To
c	14	45.6	3.9	1101	10	CNS000MZ	AL063247 Drosophil
c	15	45.2	3.9	652	3	BJ405168	BJ405168 BJ405168
c	16	45.2	3.9	1359	8	DN685273	DN685273 CGX40-G12
c	17	44.4	3.8	556	3	BI501848	BI501848 rm05h04.y
	18	44.4	3.8	870	10	CW940710	CW940710 TcB21.3_H
	19	44.4	3.8	933	10	CG768400	CG768400 TcB50.4_E
	20	44.2	3.8	1101	10	CNS003BA	AL064088 Drosophil
c	21	44.2	3.8	1101	10	CNS00FNX	AL070827 Drosophil
	22	44	3.8	563	1	AU261284	AU261284 AU261284
c	23	44	3.8	1021	10	CNS014DY	AL104032 Drosophil
c	24	43.8	3.8	726	9	BZ957657	BZ957657 PgH006M17
	25	43.8	3.8	786	7	CN205012	CN205012 Tor5403 G
	26	43.8	3.8	838	10	CZ547476	CZ547476 SRAA-aad7
	27	43.6	3.8	280	8	DN340146	DN340146 LIB3523-0
c	28	43.6	3.8	784	7	CJ386822	CJ386822 CJ386822
c	29	43.6	3.8	797	10	CNS003F8	AL064634 Drosophil
c	30	43.4	3.7	773	10	CG369547	CG369547 OG0FJ46TV
	31	43.4	3.7	897	10	CG369534	CG369534 OG0FJ46TH
	32	43.2	3.7	379	7	CK233784	CK233784 re19e01.y
	33	43.2	3.7	646	3	BI772948	BI772948 kq09h09.y
	34	43.2	3.7	657	10	CL341938	CL341938 RPCI44_26
	35	43.2	3.7	838	10	CZ527582	CZ527582 SRAA-aac4
c	36	43.2	3.7	849	8	DR044043	DR044043 NLE000021
c	37	43.2	3.7	898	10	CL122769	CL122769 ISB1-82K2
	38	43.2	3.7	1101	10	CNS0182P	AL108811 Drosophil
	39	43	3.7	1101	10	CNS002M0	AL062875 Drosophil
	40	42.8	3.7	608	5	BW544724	BW544724 BW544724
c	41	42.8	3.7	791	6	CD357360	CD357360 AGENCOURT
c	42	42.8	3.7	838	10	CZ532473	CZ532473 SRAA-aac7
c	43	42.8	3.7	852	9	AZ668752	AZ668752 ENTID01TF
	44	42.8	3.7	883	9	AZ670074	AZ670074 ENTLW29TR
	45	42.8	3.7	925	7	CN384881	CN384881 LE2TR02C0

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

```
  Query Match             5.1%;  Score 58.6;  DB 10;  Length 1101;
  Best Local Similarity   17.1%;  Pred. No. 0.003;
  Matches  120;  Conservative  287;  Mismatches  292;  Indels    4;  Gaps    1;

Qy        452 ACAAAATACGAATTTGGGATTCTCGAACCGCAAATGGGGTTGATCATTTTATAGCTAATT 511
              |  ::  ::  :::| ::: ::| :  ::  : :::   |: :::  ::    ::  |:::: 
Db       1099 ARRWGDDTWDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWW 1040

Qy        512 CTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGGCTTCAGTTATATATC 571
              : :||:: : :    :| ||:|: |: :    : |:: | ::|  :  ::   :|: 
Db       1039 WWWWATWDTWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAG 980

Qy        572 CGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAGACTATTATTTCACAG 631
              :   :::::::::: ::::: |  ::|| || :   ::: :: : ::::  : : |: 
Db        979 RRDGGRKRKDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKAD 920

Qy        632 CATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAAGCCTTTAGTAAAATGC 691
              ::   :::::|:    ::  :  :::::::: ::::: :  : :  : :| |:::::| 
Db        919 DDDGAGDKDDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGW 860

Qy        692 CGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAAAAAATAAAGAGCAAGG 751
              : :::  :|| :: :|:: ::   ::::: :  | : |::::: ::|  : ||: :::: 
Db        859 ADADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDR 800

Qy        752 CTACAGACAATATAAAATTGCTCGGTTATCAATCTTTTCCTGTTTTAAAAGAGTATATGC 811
              | ::: ||:: ::|||:  :   ||:: ::: ::|:  |::  :: :::::::::: 
Db        799 KRADDKRDAADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRA 740

Qy        812 AGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAG 871
              || :  :  : ::: ::| ::: :  : |:::::  ::| :|   :    :  :::: 
Db        739 GTAGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDR 680

Qy        872 CTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTC 931
              | ::  :  :  :   : :|:::: :  ::||:::: :  ::: :  ::  
Db        679 ATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAK 620

Qy        932 GCCCACTAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTT 991
                  :: :::| :   :  :|  :      :   :: :|| |:::: |:| ||  | 
Db        619 WDWKTRADRWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTTAT 560

Qy        992 TGCATGAAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTA 1051
              :  :  :|  : ::|      |:   :   :|   |  ||||:|  |  : |   |  :| 
Db        559 WTTWAARAAWWAWWAWATTTATWTTTWTTTWTTTTTTTTTTTTWTTTAAWWWAAWWTATWA 500

Qy       1052 GAAAAAATGCAGAAAAATTTTCTCGATCAAGATTTGAACAAGAATTTA----AGAACTTT 1107
              : |||:   | ||:|||||| |   |  : :|   :: |: : |||:   : || |||
Db        499 AWTAAAWAAAAAAAWAATTTTTTTTTTTTWAWTAAAWWTAWTWTTTTWTTTTWWAATTTT 440

Qy       1108 GTTAATGAAAAGTGGAATCTTTTCAAAACAGAACAGATTATTA 1150
              |: :| ::|| |    | |||: |||| | |: : :|:  ||
Db        439 TTWTWTTWWAATTATTTTWTTTWTAAAAAAAAWWWWWTWTATA 397
```

```
RESULT 2
CNS014PQ
LOCUS       CNS014PQ                 987 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN12P22 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL104456
VERSION     AL104456.1  GI:5616067
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .987
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN12P22"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN
```

```
  Query Match             4.3%;  Score 49.4;  DB 10;  Length 987;
  Best Local Similarity   31.9%;  Pred. No. 0.52;
  Matches  102;  Conservative   64;  Mismatches  154;  Indels    0;  Gaps    0;

Qy        515 AATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGGCTTCAGTTATATATCCGC 574
              :| : |::     :| ||||| || |  | | | |     |       | :|||::|: 
Db        642 MAAMAAWMAWWACWAAAATTAGAATAAAAAAAWAWAAAAAAAAAAATAAWTATWWAWNTNW 701

Qy        575 CTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAGACTATTATTTCACAGCAT 634
              : |: :: :    |  |||| ||  |||::|  :|||  :  | :||::||| |    |: 
Db        702 TWTTWWWWTWTTTTTTTTTTTAAWAAAAWWAAAWAAAWTWTTAWWATWWTTTAAAGAAAW 761

Qy        635 CCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAAGCCTTTAGTAAAATGCCGG 694
                ||   |:   |    :  |: | |  | |:|    :|    ||: |:::|
Db        762 TTATTAAATTWAAATTTTWWTTWATTAATTAAWTAATTWATTWATTWAWWAATTAAT 821

Qy        695 AAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAAAAAATAAAGAGCAAGGCTA 754
              |: :  |  :: |: || ||  | ::           | ::|:|||||::|:  :: 
Db        822 AWTWWTANAKWTTWWTTTTTTTTTWWTATAAATTWTTWWWAWAAATAWRGRWTWWAWAAT 881

Qy        755 CAGACAATATAAAATTGCTCGGTTATCAATCTTTTCCTGTTTTAAAAGAGTATATGCAGA 814
              | | :||::||||:    : ::  |: ::: : :|    :|| |||| |  | |:  | | 
Db        882 WAAAWWATWWAAAAWAWWWWDRAAAWDWWWWWAWTAAAWWTTAAAAAAAAAAAAWAAAAA 941

Qy        815 GCGCCAGGGCGTTTGTTTTT 834
                  :   ::| |||||
Db        942 AWWWWWWWWWWWWTTTTTTT 961
```

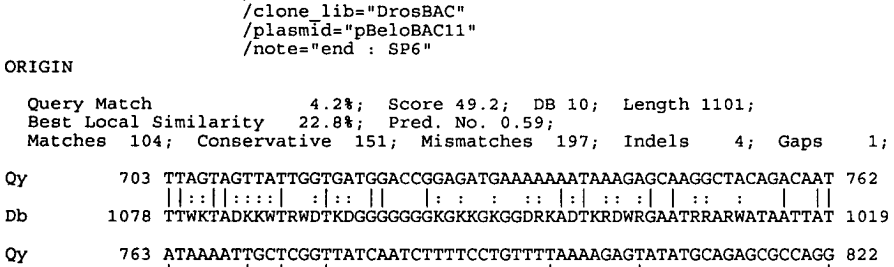
```
RESULT 3
CNS017KX/c
LOCUS       CNS017KX                 1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37F10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108171
VERSION     AL108171.1  GI:5628475
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
```

project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37F10"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : SP6"

ORIGIN

Query Match 4.2%; Score 49.2; DB 10; Length 1101;
Best Local Similarity 22.8%; Pred. No. 0.59;
Matches 104; Conservative 151; Mismatches 197; Indels 4; Gaps 1;

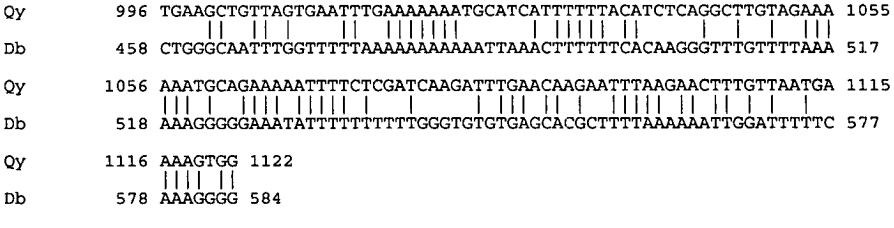


RESULT 4
BU747088
LOCUS BU747088 825 bp mRNA linear EST 10-OCT-2002
DEFINITION CH3#009_C08T7 Canine heart normalized cDNA Library in pBluescript Canis familiaris cDNA clone CH3#009_C08 5', mRNA sequence.
ACCESSION BU747088
VERSION BU747088.1 GI:23697666
KEYWORDS EST.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 825)
AUTHORS Yi,Y., Desai,R., Olarte,M., Henthorn,P. and George A.L.
TITLE Expressed sequence tags from Canine heart
JOURNAL Unpublished (2003)
COMMENT Other_ESTs: CH3#009_C08T3
Contact: George AL
Division of Genetic Medicine
Vanderbilt University
529 Light Hall, 2215 Garland Avenue, Nashville, TN 37232-0275, USA
Tel: 615 936 2660
Fax: 615 936 2661
Email: al.george@vanderbilt.edu
Insert Length: 1734 Std Error: 0.00
Seq primer: T7: TAATACGACTCACTATAGGG
High quality sequence start: 42
High quality sequence stop: 172.

FEATURES Location/Qualifiers
source 1. .825
/organism="Canis familiaris"
/mol_type="mRNA"
/db_xref="taxon:9615"
/clone="CH3#009_C08"
/tissue_type="heart"
/cell_type="heart"
/dev_stage="mixed developmental stages (adult, 30 day - 40 day fetal)"
/clone_lib="Canine heart normalized cDNA Library in pBluescript"
/note="Organ: heart; Vector: pBluescript; Site_1: 5' of vector NotI; Site_2: 3' of vector EcoRI; Tissue source: dog heart (adult, 30 day - 40 day fetal), right and left atria and ventricle. Dog breed - mixed (beagle, German shepherd, pointer, Irish setter). Library construction: oligo-dT primed"

ORIGIN

Query Match 4.1%; Score 47.8; DB 5; Length 825;
Best Local Similarity 53.5%; Pred. No. 1.2;
Matches 100; Conservative 0; Mismatches 87; Indels 0; Gaps 0;



RESULT 5
CV966702
LOCUS CV966702 872 bp mRNA linear EST 25-JAN-2005
DEFINITION PC011E12 infected tomato, center of lesion 3 dpi Phytophthora infestans cDNA, mRNA sequence.
ACCESSION CV966702
VERSION CV966702.1 GI:58156451
KEYWORDS EST.
SOURCE Phytophthora infestans (potato late blight agent)
ORGANISM Phytophthora infestans
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.
REFERENCE 1 (bases 1 to 872)
AUTHORS Randall,T., Dwyer,R.A., Huitema,E., Beyer,K., Cvitanich,C., Kelkar,H., Fong,A.M., Gates,K., Roberts,S., Yatzkan,E., Gaffney,T., Law,M., Testa,A., Torto-Alalibo,A., Zhang,M., Zheng,L., Mueller,E., Windass,J., Binder,A., Birch,P.R.J., Gisi,U., Govers,F., Gow,N.A., Mauch,F., van West,P., Waugh,M.E., Yu,J., Boller,T., Kamoun,S., Lam,S.T. and Judelson, H.S.
TITLE Large-scale gene discovery in the oomycete Phytophthora infestans reveals likely components of phytopathogenicity shared with true fungi
JOURNAL Mol. Plant-Microbe Interact. 18 (3), 229-243 (2005)
PUBMED 15782637
COMMENT Contact: Judelson HS

```
           Department of Plant Pathology
           University of California
           Webber Hall, Riverside, CA 92521, USA
           Tel: 909 787 4199
           Fax: 909 787 4294
           Email: howard.judelson@ucr.edu
           may be of host plant origin.
FEATURES             Location/Qualifiers
     source          1. .872
                     /organism="Phytophthora infestans"
                     /mol_type="mRNA"
                     /strain="88069"
                     /db_xref="taxon:4787"
                     /sex="A1"
                     /clone_lib="infected tomato, center of lesion 3 dpi"
                     /note="Vector: pSPORT1"
ORIGIN

  Query Match             4.1%;  Score 47.6;  DB 8;  Length 872;
  Best Local Similarity   48.2%;  Pred. No. 1.4;
  Matches  134;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy        188 AAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCTA 247
              ||| ||  |||| |  | ||  ||| |  ||       || | ||   ||| | |   ||
Db         46 AAACTTATCCTACACTTTAACAATATCTCCTTTTGCCTTGAACAACTATGGCTCCTTCTC 105

Qy        248 TTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCAA 307
              || ||||||    |  ||| || || |  | || |  |       |  |          |
Db        106 TTCAACAACCAATTCCATCTGAAGCAATAACCAGCGATGCATCATCGGACGTTACCATCA 165

Qy        308 AAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTATTCGAT 367
               ||    || | |  || || ||  ||     |  | | || || ||||| |||
Db        166 CTGGAAAGATATACACACGAGTTCGTCTCGCTACGAAATCTGATCTTTCTCATATATACC 225

Qy        368 ATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAAG 427
               ||   |   |   ||| | |||| |||  |  || | ||  | ||||  | | |
Db        226 GATTGTTTTACCAAATCCATGAATACCATAACTATACTCATCTATACAAAGCTACTGAGT 285

Qy        428 GTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATT 465
               |  || || || | |||  || |  |||  |   ||
Db        286 CCTCCTTGGCCAACTTGCTCTTTAAGGAAAACCCTCTT 323


RESULT 6
BU279377
LOCUS       BU279377                872 bp    mRNA    linear   EST 27-NOV-2002
DEFINITION  603864720F1 CSEQCHN54 Gallus gallus cDNA clone ChEST881h5 5', mRNA
            sequence.
ACCESSION   BU279377
VERSION     BU279377.1  GI:25728831
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 872)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .872
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Compton Line 15I"
                     /db_xref="taxon:9031"
                     /clone="ChEST881h5"
                     /sex="Female"
                     /tissue_type="not cerebrum or cerebellum"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN54"
                     /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: Not1; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
ORIGIN

  Query Match             4.0%;  Score 46.6;  DB 5;  Length 872;
  Best Local Similarity   48.7%;  Pred. No. 2.5;
  Matches  127;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy        483 AAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGT 542
              ||||    ||||||| ||| | |  || | |  | |  |  |   | ||||| |
Db         63 AAATAACTTTGATCACTTTGTTGAGAAGTTTTCAAAGGTGCCATCTTAAATTACATCCAA 122

Qy        543 ATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAA 602
              |||  |  ||  |   || | |  | || |         |||    |||||| |  ||||
Db        123 ATAATGTTGTATGCTATCTGCTTGAGCTCAGTACTATTCTGTACTGAATTTTAACCTAAA 182

Qy        603 AAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGA 662
               | ||   | || |  || | | |   |||   |  |  || |  ||||  | | ||||
Db        183 CAGTGCCTTGAAACAAAATAAATGCGATGCAGAACTAAGAGTTCTGTACATTCCTTTTGA 242

Qy        663 TCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGG 722
              | || ||   |  | | ||  | |  |   ||    ||     | ||| || |   |||
Db        243 TATTTTTCATAATTCTTGTAAAACATCGGGAGAGGCCAAGCACATTGTTTTTTGCATTGG 302

Qy        723 ACCGGAGATGAAAAAAATAAA 743
                |||   || |||  | |
Db        303 TGATGAGGCAAATAAAGAACA 323


RESULT 7
CNS02Q6S
LOCUS       CNS02Q6S                924 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            157A04 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL208909
VERSION     AL208909.1  GI:7867728
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
   PUBMED   10835645
```

```
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
   PUBMED   10899143
REFERENCE   3  (bases 1 to 924)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .924
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="157A04"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG157BA02SP1
                     end : PUC-Ori"
ORIGIN

  Query Match             4.0%;  Score 46.6;  DB 10;  Length 924;
  Best Local Similarity   47.6%;  Pred. No. 2.5;
  Matches   70;  Conservative   18;  Mismatches   59;  Indels    0;  Gaps    0;

Qy       1012 TTTGAAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTT 1071
                |:| ||||:|: :   |  ||||||   | : |   :| ||  |:|||      :||| |:
Db        651 TWTAAAAAWATWWTTTTTTTTTTTTTNTTWAAATTWTTTATTAWAAAAAATTTWAAAATW 710

Qy       1072 TCTCGATCAAGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTC 1131
                : |  ||  | : |: || ||   |: || :  ||| |: |  |||| |  : | |||:
Db        711 WATATATWTAAWATWTAAAAAATTTWAAAAWTTTTTWTWTAAAAAAAATTWWTTWTTTWW 770

Qy       1132 AAAACAGAACAGATTATTAAACGTTAA 1158
                |||| | || | | :|| |||  |:||
Db        771 AAAAAAAAAWAWANWATAAAAWATWAA 797


RESULT 8
BJ407454/c
LOCUS       BJ407454                582 bp    mRNA    linear   EST 10-MAR-2002
DEFINITION  BJ407454 Dictyostelium discoideum cDNA library, SF Dictyostelium
            discoideum cDNA clone dds38i18 3', mRNA sequence.
ACCESSION   BJ407454
VERSION     BJ407454.1  GI:19320346
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the slug stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dds38i18"
                     /sex="mat A"
                     /dev_stage="Slug stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN

  Query Match             4.0%;  Score 46.2;  DB 3;  Length 582;
  Best Local Similarity   52.0%;  Pred. No. 2.9;
  Matches  102;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        535 AAAAAAGTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTT 594
              ||| || || |   |  ||| ||| | | |     |  |  ||  ||||  |||||
Db        553 AAACAATTAAAACAAGATGATGCTGCCGCTGCTGCTGAGGATGATGATGAAAATAATGAA 494

Qy        595 GAAGTAAAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAA 654
              || |   || |  || | ||| |  |   |   | | ||    ||     ||   | |||
Db        493 GATGATGAAGAACAACAACAACAAGAAGTTAAAAAACCAAAGGGTGNTAAACAAAAGAAA 434

Qy        655 CGTATTGATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATT 714
              |  |||  || | | || |||      |   ||| ||  ||||||||||||   || |
Db        433 CCAATTACTCCTGTCGTAGAAAAAGAAATCGAAAAGCTAGAAAAGAAATTAGAGATTTTA 374

Qy        715 GGTGATGGACCGGAGA 730
              | |||      |||||
Db        373 GATGAAAAGTTGGAGA 358


RESULT 9
CNS0039G
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscopé - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN
```

```
  Query Match             4.0%;  Score 46.2;  DB 10;  Length 1101;
  Best Local Similarity   14.2%;  Pred. No. 3.2;
  Matches   78;  Conservative  239;  Mismatches  232;  Indels    0;  Gaps    0;

Qy        166 TATGCGACTACCACATTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATAT 225
              |:|    : |:  | :||:::| |||:   |   |||   ||:|||: | |  : :: ::|
Db        492 TWTTTAWTTWATAWWTTWWWTTAAAWAAAAAAAAAAAAWAAAWAAAWATAAATWTWWTWWT 551

Qy        226 TTACCACTAATGCCACTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATT 285
              |:   | :|    : :   :::|:        :::| :|:|:| : :| :|:|:|:  :::::
Db        552 TYTTWAAWATAAAMCMAAWYYHTYTTYTYHYYTYWTYTMTWHYHTMYTHAWAHTTWYHWY 611

Qy        286 AGTAGCGCCCATTCCGTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGC 345
              : :|    : ::: :  : ::        :: :| ::  :::: :: : :   :  |:
Db        612 HTYAMWHWMTWHWTMWAWWHWTTYTAAYYYYYTCMYYYHYMHWHHAHAHAAAWWTTHTWW 671

Qy        346 TATGTTCATTCTCCTATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCT 405
              |:  ::  ::::::    :    |   :    :::::  ::|:  ::: :: ::   : :
Db        672 THAYHWATYHYYYYMYCAMMCMCTHTCHHCYYYYHHYTAHHTHTHHWYAHYYMWYYMWAY 731

Qy        406 AACCTGAATAAAGGAATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATT 465
              :::  | | :   ::::::|  ||   :|  :::::  ::      :  |:| |:: ::||
Db        732 YWMYCTACTYHYHHHHHYHWAYHTTWYAWAHAMWMWHHAHYAAAAAWAAWATTHHYHHTT 791

Qy        466 TGGGATTCTCGAACCGCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCG 525
              :   :|::::  ::  | : |    :: : :|  | :|:|:   : :|:::::|::::
Db        792 HYMHHTYMYHYYMYTCCYMCTYHCWHYYHTAYTCWTWTHHWMWTWTHWYHHTWWHHTTTH 851

Qy        526 CGTAGAATTAAAAAAGTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTG 585
              ::  :|: ::::: | :  | :|   :  ::::: : ::::: |: ::: :: :  :
Db        852 WAWWHTHTWCWWWWHATTWTWATHCWACMTMHWHHWMHMHHHHMACHAHHHTHMCMCHHH 911

Qy        586 GATAATTTTGAAGTAAAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTA 645
              :  :::: :     ::::: :::: ::: | ::||:|| :::  ::::::|   :  ::
Db        912 MHCTCHHHHTMYHMTCHWMWHMHWHWHHWWMATWMTTMTTMMMMCCMMHHHCHMYHMMHM 971

Qy        646 CCCTACAAACGTATTGATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTA 705
              :::  :     : :  ::::: | ::    : ::|| ::|:   :  | :: :::::|
Db        972 YMYCCHYYCTCHTHATTHYHYMCTCYHYCTWHTYWTAYWWAWTAHAMTTATWWWWMHWWA 1031

Qy        706 GTAGTTATT 714
              :| :::::
Db       1032 HWATWWWWW 1040
```

```
RESULT 10
BW542444
LOCUS       BW542444                498 bp    mRNA    linear   EST 31-AUG-2004
DEFINITION  BW542444 Yutaka Satou unpublished cDNA library (cscl) Ciona
            savignyi cDNA clone cscl001h07 3', mRNA sequence.
ACCESSION   BW542444
VERSION     BW542444.1  GI:51722316
KEYWORDS    EST.
SOURCE      Ciona savignyi
  ORGANISM  Ciona savignyi
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Vinson,J.P., Jaffe,D.B., O'neill,K., Karlsson,E.K.,
            Stange-Thomann,N., Anderson,S., Mesirov,J.P., Satoh,N., Satou,Y.,
            Nusbaum,C., Birren,B., Galagan,J.E. and Lander,E.S.
  TITLE     Assembly of polymorphic genomes: Algorithms and application to
            Ciona savignyi
  JOURNAL   Genome Res. 15 (8), 1127-1135 (2005)
   PUBMED   16077012
COMMENT     Contact: Yutaka Satou
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4095
            Fax: 81-75-705-1113
            Email: yutaka@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Ciona savignyi"
                     /mol_type="mRNA"
                     /db_xref="taxon:51511"
                     /clone="cscl001h07"
                     /dev_stage="cleavage stage"
                     /clone_lib="Yutaka Satou unpublished cDNA library (cscl)"
ORIGIN
```

```
  Query Match             4.0%;  Score 46;  DB 5;  Length 498;
  Best Local Similarity   56.3%;  Pred. No. 3.2;
  Matches   85;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy        964 TTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAAAAT 1023
              ||   | || || | || |||| | ||| ||| ||     |||||||| |   | | |
Db          9 TTTAACCAGNAATAAAAAATTGAATATTTTCATAAATGATTTAGTGAAACTATTATTATT 68

Qy       1024 GCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCAAGA 1083
              ||| ||||  | | ||   | ||  | || |||  | || |  |||   | ||  |
Db         69 TTATCGTTTTGAATAACTGGTGTTTTGAAAAGAAATCAAAAATACATTTAAAGCACACTA 128

Qy       1084 TTTGAACAAGAATTTAAGAACTTTGTTAATG 1114
              | |  |  ||| |   |||| |||||||
Db        129 CATTATAATAAATATGCAAACTAAGTTAATG 159
```

```
RESULT 11
BJ348545/c
LOCUS       BJ348545                635 bp    mRNA    linear   EST 06-MAR-2002
DEFINITION  BJ348545 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda33f18 3', mRNA sequence.
ACCESSION   BJ348545
VERSION     BJ348545.1  GI:19219052
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .635
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda33f18"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN
```

```
  Query Match             4.0%;  Score 46;  DB 3;  Length 635;
  Best Local Similarity   51.0%;  Pred. No. 3.3;
  Matches  100;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        535 AAAAAAGTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTT 594
              |||  | || |   |  ||| ||| | | |      |  |  ||  ||||  |||||
Db        569 AAACNATTAAAACAAGATGATGCTGCCGCTGCTGCTGAGGATGATGATGAAAATAATGAA 510

Qy        595 GAAGTAAAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAA 654
              || |   || |  ||  ||| | |   |  | | ||   ||    ||   | |||
Db        509 GATGATGAAGAACAACNACAACAAGAAGTTAAAAAACCAAAGGGTGNTAAACAAAAGAAA 450
```

```
Qy        655 CGTATTGATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATT 714
              |  |||  || | | || |||      |   ||| ||  ||||||||||||   || |
Db        449 CCNATTACTCCTGTNGTAGAAAAAGAAATNGAAAAGCTAGAAAAGAAATTAGAGATTTTA 390

Qy        715 GGTGATGGACCGGAGA 730
              | |||      |||||
Db        389 GATGAAAAGTTGGAGA 374


RESULT 12
BJ356306/c
LOCUS       BJ356306                678 bp    mRNA    linear   EST 07-MAR-2002
DEFINITION  BJ356306 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda59123 3', mRNA sequence.
ACCESSION   BJ356306
VERSION     BJ356306.1  GI:19255901
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .678
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda59123"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

  Query Match             4.0%;  Score 46;  DB 3;  Length 678;
  Best Local Similarity   48.8%;  Pred. No. 3.3;
  Matches  118;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy        489 GGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAG 548
              || | || |  ||| |     |      ||  ||||           ||| || || |
Db        613 GGGTAATGAAGTTAAACAAGAGAAAAAAGGTATCATNTTAGAGGAAAAACAATTAAAACA 554

Qy        549 ACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGA 608
              |  ||| ||| | | |     |  |  ||  ||||   |||||   || |   || |  |
Db        553 AGATGATGCTGCCGCTGCTGCTGAGGATGATGATGAAAATAATGAAGATGATGAAGAACA 494

Qy        609 AAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTAT 668
              |   ||| |  |   |   | | ||    ||     ||   | ||||  |||  || | |
Db        493 ACNACAACAAGAAGTTAAAAAACCAAAGGGTGNTAAACAAAAGAAACCAATTACTCCTGT 434

Qy        669 TGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGA 728
              || |||       |   ||| ||  ||||||||||||   || | | |||      |||
Db        433 CGTAGAAAAAGAAATNGAAAAGCTAGAAAAGAAATTAGAGATTTTAGATGAAAAGTTGGA 374

Qy        729 GA 730
              ||
Db        373 GA 372


RESULT 13
CZ974493/c
LOCUS       CZ974493                930 bp    DNA     linear   GSS 11-AUG-2005
DEFINITION  175383 Tomato MboI BAC Library Lycopersicon esculentum genomic
            clone SL_MboI0102L08 5, genomic survey sequence.
ACCESSION   CZ974493
VERSION     CZ974493.1  GI:72326138
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 930)
  AUTHORS   Mueller,L.A., Buels,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J.,
            Van Eck,J. and Stack,S.
  TITLE     BAC end sequencing from three Solanum lycopersicon libraries
  JOURNAL   Unpublished (2005)
COMMENT     Other_GSSs: 216382
            Contact: Lukas Mueller
            Tanksley Lab, Dept. of Plant Breeding
            Cornell University
            251 Emerson Hall, Ithaca, NY 14853, USA
            Tel: 607-255-6557
            Fax: 607-255-6683
            Email: sgn-feedback@sgn.cornell.edu
            Plate: 102  row: L  column: 8
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 8
            High quality sequence stop: 581.
FEATURES             Location/Qualifiers
     source          1. .930
                     /organism="Lycopersicon esculentum"
                     /mol_type="genomic DNA"
                     /cultivar="Heinz 1706"
                     /db_xref="taxon:4081"
                     /clone="SL_MboI0102L08"
                     /lab_host="E. coli"
                     /clone_lib="Tomato MboI BAC Library"
                     /note="Vector: pBeloBAC11; Site_1: MboI"
ORIGIN

  Query Match             4.0%;  Score 45.8;  DB 10;  Length 930;
  Best Local Similarity   52.3%;  Pred. No. 3.9;
  Matches  101;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        960 CATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAA 1019
              | | |  |||||  |  |   |||| |||   |  | ||  | ||| || |  | |||
Db        289 CCTATCTTTCAATTAGTAATTTATTTCTTAAGTATTTAAAATTTTTAATGGAAATCAAAT 230

Qy       1020 AAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATC 1079
              || ||||||| | || |  | | |        | || |   | || |||  |    ||
Db        229 CGATTCATCATTGTGTAAACATTAAGGACAATAATAATAATGATAATAATAATGATAATG 170

Qy       1080 AAGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGA 1139
              |||   |||||  ||   | ||     | ||| | || | ||   |||  | || |||
Db        169 TCGATGATAACAATTATAATAAAAAGCCATCAATTAGAATTAGAGCATTTATAGAAAAGA 110

Qy       1140 ACAGATTATTAAA 1152
              |  | ||||||
Db        109 CAATCTAATTAAA 97


RESULT 14
CNS000MZ/c
LOCUS       CNS000MZ                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR02G15 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063247
VERSION     AL063247.1  GI:4938983
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

```
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR02G15"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             3.9%;  Score 45.6;  DB 10;  Length 1101;
  Best Local Similarity   38.3%;  Pred. No. 4.5;
  Matches   80;  Conservative   39;  Mismatches   90;  Indels    0;  Gaps    0;

Qy        941 GTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAG 1000
              | | :: || | ||:| | |::||:| | | ||::|| |||  :
Db       1054 GCGGRSKGGKAAAAACKGAWAAASGAARKGADRAAMAMATTTTTBYTTWTTTTTTWTTTW 995

Qy       1001 CTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATG 1060
              |  ||    || ||  |:|:       |::::|  || | |  :|| | ||||| |
Db        994 CAWTTTAAAAAATTTYTAWAWTAATTAAGTKWYKTTWATWTWAAAYTTAAAAAAAATTA 935

Qy       1061 CAGAAAAATTTTCTCGATCAAGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGT 1120
                :|| :||||: : | |: | :||  |  || :||:|    ||:|  : :| : | :::
Db        934 TTDAATWATTTKTWAGTTSGAAWTTATAYTAAAWATKTTTTTACWTCTYMMAARGANRBB 875

Qy       1121 GGAATCTTTTCAAAACAGAACAGATTATT 1149
              :|: :::  ::|||: | |  | ||| ||
Db        874 BGVBBVBSMHMAAAWAAWATAATATTTTT 846


RESULT 15
BJ405168/c
LOCUS       BJ405168                652 bp    mRNA    linear   EST 10-MAR-2002
DEFINITION  BJ405168 Dictyostelium discoideum cDNA library, SF Dictyostelium
            discoideum cDNA clone dds30f19 3', mRNA sequence.
ACCESSION   BJ405168
VERSION     BJ405168.1  GI:19318085
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the slug stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .652
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dds30f19"
                     /sex="mat A"
                     /dev_stage="Slug stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN

  Query Match             3.9%;  Score 45.2;  DB 3;  Length 652;
  Best Local Similarity   47.9%;  Pred. No. 5.2;
  Matches  116;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy        489 GGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAG 548
              || | || |  ||| |     |      ||  ||||             |||  | || |
Db        613 GGGTAATGAAGTTAAACAAGAGAAAAAAGGTATCATNTTAGAGGAAAAACNATTAAAACA 554

Qy        549 ACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGA 608
              |  ||| ||| | | |     |  |  ||  ||||   |||||   || |   || |
Db        553 AGATGATGCTGCCGCTGCTGCTGAGGATGATGATGAAAATAATGAAGATGATGAAGAACN 494

Qy        609 AAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTAT 668
              |   ||| |  |   |  | | ||    ||     ||   | ||||  |||  || | |
Db        493 ACNACAACAAGAAGTTAAAAAACCAAAGGGTGNTAAACAAAAGAAACCNATTACTCCTGT 434

Qy        669 TGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGA 728
              || |||       |   ||| ||  ||||||||||||   || | | |||       |||
Db        433 NGTAGAAAAAGAAATNGAAAAGCTAGAAAAGAAATTAGAGATTTTAGATGAAAAGTTGGA 374

Qy        729 GA 730
              ||
Db        373 GA 372


Search completed: May 12, 2006, 03:29:55
Job time : 3616.95 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 19:15:19 ; Search time 158.379 Seconds
(without alignments)
12996.800 Million cell updates/sec

Title: US-09-423-093B-3_COPY_7310_8467
Perfect score: 1158
Sequence: 1 atggaaataaatgaaataaa.....aacagattattaaacgttaa 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	520	44.9	1167	3	US-09-489-039A-999	Sequence 999, App
c	2	181.4	15.7	324	3	US-09-489-039A-1205	Sequence 1205, Ap
c	3	120.2	10.4	222	3	US-09-489-039A-1126	Sequence 1126, Ap
	4	59.8	5.2	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	5	47.6	4.1	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	6	43.8	3.8	7218	2	US-08-232-463-14	Sequence 14, Appl
	7	42.2	3.6	456	3	US-09-902-540-1853	Sequence 1853, Ap
c	8	42.2	3.6	5752	3	US-09-902-540-746	Sequence 746, App
	9	41	3.5	601	3	US-09-949-016-203910	Sequence 203910,
	10	41	3.5	194537	3	US-09-949-016-12928	Sequence 12928, A
	11	41	3.5	201529	3	US-09-949-016-12740	Sequence 12740, A
c	12	40	3.5	421491	3	US-09-949-016-12805	Sequence 12805, A
c	13	40	3.5	421494	3	US-09-949-016-14060	Sequence 14060, A
c	14	39.8	3.4	38119	3	US-09-949-016-17503	Sequence 17503, A
c	15	39.4	3.4	7175	3	US-09-722-377-9	Sequence 9, Appli
c	16	39.4	3.4	317366	3	US-09-949-016-16001	Sequence 16001, A
	17	39.2	3.4	121384	3	US-09-949-016-16944	Sequence 16944, A
c	18	39	3.4	1689	2	US-07-991-867B-41	Sequence 41, Appl
c	19	39	3.4	1689	2	US-08-544-332-41	Sequence 41, Appl
c	20	39	3.4	1689	3	US-09-370-861A-41	Sequence 41, Appl
c	21	39	3.4	1947	3	US-09-370-861A-74	Sequence 74, Appl
	22	39	3.4	1947	3	US-09-662-254B-55	Sequence 55, Appl
c	23	39	3.4	8457	2	US-07-991-867B-1	Sequence 1, Appli
c	24	39	3.4	8457	2	US-08-544-332-1	Sequence 1, Appli
c	25	39	3.4	8457	3	US-09-370-861A-1	Sequence 1, Appli
c	26	39	3.4	50000	3	US-09-662-254B-26	Sequence 26, Appl
c	27	38.8	3.4	5397	3	US-09-949-016-14334	Sequence 14334, A
c	28	38.8	3.4	19932	2	US-08-477-451-25	Sequence 25, Appl
c	29	38.6	3.3	183	3	US-09-489-039A-952	Sequence 952, App
c	30	38.6	3.3	727	3	US-09-270-767-7835	Sequence 7835, Ap
c	31	38.6	3.3	727	3	US-09-270-767-23117	Sequence 23117, A
c	32	38.6	3.3	50000	3	US-09-662-254B-23	Sequence 23, Appl
c	33	38.4	3.3	2977	3	US-09-023-655-1042	Sequence 1042, Ap
c	34	38.4	3.3	3230	6	PCT-US93-06251-89	Sequence 89, Appl
c	35	38.4	3.3	90876	3	US-09-949-016-13271	Sequence 13271, A
	36	38.4	3.3	640681	3	US-09-790-988-1	Sequence 1, Appli
c	37	38.2	3.3	351	3	US-09-134-001C-2295	Sequence 2295, Ap
c	38	38	3.3	601	3	US-09-949-016-120408	Sequence 120408,
	39	38	3.3	219964	3	US-09-949-016-15086	Sequence 15086, A
	40	38	3.3	1664976	3	US-08-916-421B-1	Sequence 1, Appli
	41	38	3.3	1664976	3	US-09-692-570-1	Sequence 1, Appli
c	42	37.8	3.3	1230025	3	US-09-198-452A-1	Sequence 1, Appli
c	43	37.8	3.3	1230230	3	US-09-438-185A-1	Sequence 1, Appli
c	44	37.4	3.2	1035	3	US-09-248-796A-1843	Sequence 1843, Ap
	45	37.2	3.2	1071	3	US-09-248-796A-7112	Sequence 7112, Ap

ALIGNMENTS

RESULT 1
US-09-489-039A-999
; Sequence 999, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 999
; LENGTH: 1167
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-999

Query Match 44.9%; Score 520; DB 3; Length 1167;
Best Local Similarity 65.7%; Pred. No. 7.3e-127;
Matches 757; Conservative 0; Mismatches 395; Indels 0; Gaps 0;

```
Qy          7 ATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTCCGAA 66
              | ||  || || || ||   | ||||   ||| |||||  | | |||||| || || |||
Db         16 AAAACGGACATTAACATTGGTGTCGTGGCTGATTGGTTTATCACTTATGCTGGTTCAGAA 75

Qy         67 CAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGTTGAT 126
               ||||     |||   |  |   | | |||||| ||    || | ||  | || || |||
Db         76 AAGGTTGTCGCTGAGTTTTTAAGTTTATTTCCTAAATGCGAGCTTTACGCTGTTGTGGAT 135

Qy        127 TTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATTTATT 186
              |||||  |  | ||  | | |  |||||||| |   |||      || || || || |||
Db        136 TTTCTCTCTCAAGAGAATAAATTACATTTTCAGAATAAAGTAATAACAACTACTTTCATT 195

Qy        187 CAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCT 246
              ||||| || ||||||||    | |   ||||||  ||| |||||| | ||||||||||
Db        196 CAAAACTTGCCTAAAGCACGTACAAAATACCAGTCATACTTACCATTTATGCCACTGGCA 255

Qy        247 ATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCA 306
              |||||||| |||||| | ||  |   | || | ||  ||  |||   |||| | || ||
Db        256 ATTGAACAGCTTGATGTCTCTAAGCATGATGTAATACTTCTTAGTAGCCATGCAGTCGCT 315

Qy        307 AAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTATTCGA 366
              || || ||  || | ||||| || || |||||||| || ||||||||||||||||||||
Db        316 AAGGGCGTGCTTACAGGACCTGACCAACTTCACATAAGTTATGTTCATTCTCCTATTCGT 375
```

```
Qy        367 TATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAA 426
              ||||| |||||||| || ||||||||| | |  ||| ||    ||| ||| ||| | |||
Db        376 TATGCATGGGATTTGCAACATCAGTACTTGAGAGAGGCTGGATTGAGTAAGGGAGTGAAA 435

Qy        427 GGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAAT 486
              | |   || ||    ||| |  | || |||||| | || |||||||| || || ||||||
Db        436 GCTCTCTTGGCTCGGTGGATATTACATAAAATAAGGATGTGGGATTCCCGGACTGCAAAT 495

Qy        487 GGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATAC 546
              || || || ||||| || |||||||| | | ||| |||||| | || |||||||| ||
Db        496 GGTGTAGACCATTTCATTGCTAATTCTAATTTTATTGCGCGTCGTATAAAAAAAGTCTAT 555

Qy        547 AGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAAT 606
              |  |  | ||    |||||||||| ||||| ||||| || || ||   | |  || |
Db        556 GGCAGGAATGCGGATGTTATATATCCTCCTGTTGATGTCGACAACTTCACATTTCAAGAA 615

Qy        607 GAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTT 666
              || ||| |||| |||||||| || ||||| ||| |||| || || ||||||||||||| |
Db        616 GAGAAGGAAGAATATTATTTTACGGCATCTCGTTTGGTTCCGTATAAACGTATTGATTTG 675

Qy        667 ATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCG 726
              ||||| || || ||||||    ||||| | ||  ||| | || ||| | ||||| |||||
Db        676 ATTGTTGAGGCATTTAGTTGTATGCCAGGAAGAAAACTTGTTGTTGTCGGTGACGGACCC 735

Qy        727 GAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTATCAATCT 786
              || |||| ||| ||||| |  || || ||| | || ||| |  | || || || || ||
Db        736 GAAATGAGAAAGATAAAAAATAAAGCAACAAAAAACATAGATATTCTAGGCTACCAGTCA 795

Qy        787 TTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAG 846
              |   | | |    |||||||||||| |  || |  || |||||||||| || |||||
Db        796 AATGAGGCTATGCGGGAGTATATGCAAAAAGCAAAAGCATTTGTTTTTGCGGCCGAAGAA 855

Qy        847 GACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTGGG 906
              || ||||| || |  || || || || || || || || || ||||||||||||||||||
Db        856 GATTTTGGCATTACCCCGGTTGAGGCCCAGGCGTGTGGAACACCTGTTATTGCCTTTGGT 915

Qy        907 AAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGCATTTTC 966
              || |||||  | || ||||| ||  | |||   || |||    | || ||||| |||||
Db        916 AAAGGTGGATCTTTGGAAACAGTAAGACCATATGGAGTAAGCAATCCTACTGGTATTTTA 975

Qy        967 TTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAAAATGCA 1026
              ||  |  | ||  ||||    ||| |   | | | ||||  |  |||||||
Db        976 TTTGAAAACCAAGATATACAATCTATAATTAACGGTGTTGATTTATTTGAACGTGTATGC 1035

Qy       1027 TCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCAAGATTT 1086
                 |||   ||    | |||||||||  |||| | |||||||||       |||||||
Db       1036 GATGAAATTAGTCCTGGAGATTGTAGAAAGCATGCTGTAAAATTTTCCATTGATAGATTT 1095

Qy       1087 GAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAACAGATT 1146
                | ||| | ||  | | ||||    ||||| |  ||    ||| |   |   | |
Db       1096 CGTAATGAAATAAATGATTATGTTCTAGAAAAATATAACAAATTCTTAGATGCGAAAAAA 1155

Qy       1147 ATTAAACGTTAA 1158
              || ||  ||||
Db       1156 ATAAATTATTAA 1167
```

RESULT 2
US-09-489-039A-1205/c
; Sequence 1205, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1205
; LENGTH: 324
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1205

Query Match 15.7%; Score 181.4; DB 3; Length 324;
Best Local Similarity 73.0%; Pred. No. 5.2e-38;
Matches 233; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```
Qy        449 TTCACAAAATACGAATTTGGGATTCTCGAACCGCAAATGGGGTTGATCATTTTATAGCTA 508
              | || |||||| | || |||||||| || || |||||||| || || ||||| || ||||
Db        323 TACATAAAATAAGGATGTGGGATTCCCGGACTGCAAATGGTGTAGACCATTTCATTGCTA 264

Qy        509 ATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGGCTTCAGTTATAT 568
              ||||| | | ||| |||||| | || |||||||| ||  |  |  | ||     |||||||
Db        263 ATTCTAATTTTATTGCGCGTCGTATAAAAAAAGTCTATGGCAGGAATGCGGATGTTATAT 204

Qy        569 ATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAGACTATTATTTCA 628
              |||| ||||| ||||| || || ||    | |  || | || ||| |||| |||||||| |
Db        203 ATCCTCCTGTTGATGTCGACAACTTCACATTTCAAGAAGAGAAGGAAGAATATTATTTTA 144

Qy        629 CAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAAGCCTTTAGTAAAA 688
              | ||||| ||| |||| || || ||||||||||||| | ||||| || || ||||||    |
Db        143 CGGCATCTCGTTTGGTTCCGTATAAACGTATTGATTTGATTGTTGAGGCATTTAGTTGTA 84

Qy        689 TGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAAAAAATAAAGAGCA 748
              |||| | ||  ||| | || ||| | ||||| ||||| || |||| ||| |||||  |  |
Db         83 TGCCAGGAAGAAAACTTGTTGTTGTCGGTGACGGACCCGAAATGAGAAAGATAAAAAATA 24

Qy        749 AGGCTACAGACAATATAAA 767
              | || ||| | || ||| |
Db         23 AAGCAACAAAAAACATAGA 5
```

RESULT 3
US-09-489-039A-1126/c
; Sequence 1126, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1126
; LENGTH: 222
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1126

Query Match 10.4%; Score 120.2; DB 3; Length 222;
Best Local Similarity 71.5%; Pred. No. 5.6e-22;
Matches 158; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        226 TTACCACTAATGCCACTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATT 285
              |||||| | |||||||||||| |||||||| |||||| | ||  |   | || | ||  ||
Db        222 TTACCATTTATGCCACTGGCAATTGAACAGCTTGATGTCTCTAAGCATGATGTAATACTT 163

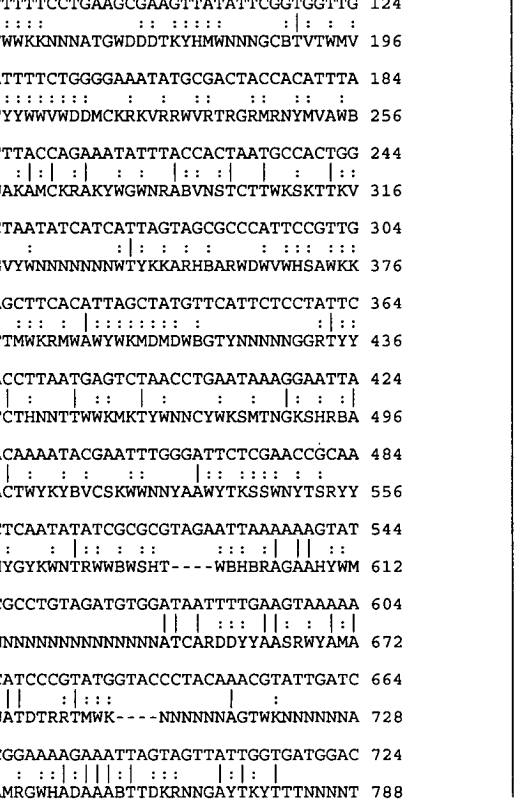
Qy        286 AGTAGCGCCCATTCCGTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGC 345
                  |||   |||| | || || || || ||  || | ||||| || || ||||||| ||
Db        162 CTTAGTAGCCATGCAGTCGCTAAGGGCGTGCTTACAGGACCTGACCAACTTCACATAAGT 103

Qy        346 TATGTTCATTCTCCTATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCT 405
              |||||||||||||||||||| ||||| |||||||| || |||||||| | |  ||| ||
Db        102 TATGTTCATTCTCCTATTCGTTATGCATGGGATTTGCAACATCAGTACTTGAGAGAGGCT 43

Qy        406 AACCTGAATAAAGGAATTAAAGGTTGGTTAGCAAAATGGCT 446
                  ||| ||| ||| | |||| |   || ||    ||| |
Db         42 GGATTGAGTAAGGGAGTGAAAGCTCTCTTGGCTCGGTGGAT 2
```

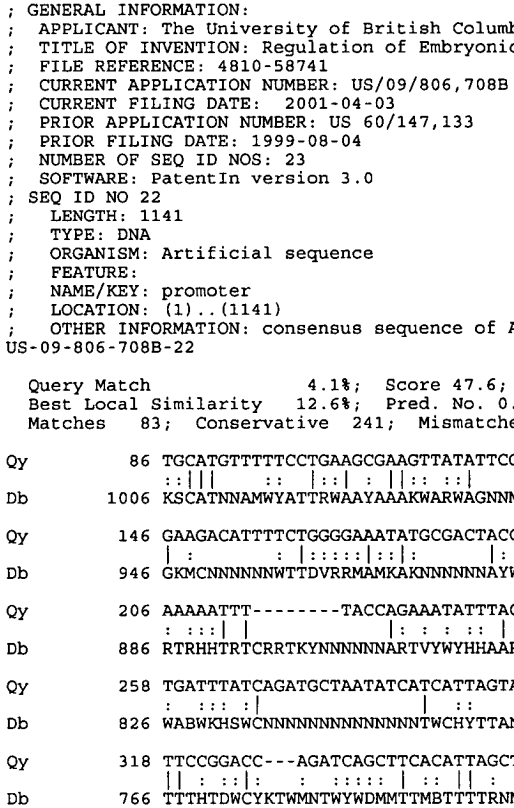
RESULT 4
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 5.2%; Score 59.8; DB 3; Length 1141;
Best Local Similarity 12.7%; Pred. No. 8e-06;
Matches 128; Conservative 359; Mismatches 502; Indels 15; Gaps 4;



RESULT 5
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 4.1%; Score 47.6; DB 3; Length 1141;
Best Local Similarity 12.6%; Pred. No. 0.013;
Matches 83; Conservative 241; Mismatches 323; Indels 14; Gaps 3;



```
Qy        375 GGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTAAAGGTTGGTT 434
Db        706 YYAHATNNWGCWWNNTDARRTNNTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNNNNNNN 647

Qy        435 AGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAAATGGGGTTGA 494
Db        646 NNNNNNNNSCCTCTRMMTMRWTMKGDGMTVRKKVKWRDTTCTYVDVWADSWVWWYANWMR 587

Qy        495 TCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGA 554
Db        586 CRDVTYTRNNTYCKSYAHSYWYWSNNAMWYRRYSARNWSSMARWTTRNNWWMSGBVRMRW 527

Qy        555 GGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCA 614
Db        526 AGTMWWRHWNNNNTDTRYYWWWKRWARBTTTVYDSMCNAKSMWRGNNWRAMKMWWAANND 467

Qy        615 AGACTATTAT---TTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGT 671
Db        466 AGAMDHWTYWMGNNTMWMRRAWKMMNMAWCRRAYCCNNNNNRACVWHKHKMWRWTWKYMW 407

Qy        672 CGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGAT 731
Db        406 KAACNNNNBKAMYMRVAWMMYSRDTTNTDWMMWTSDWBWHWYTVDYTMMRAWNNNNNNNW 347

Qy        732 G 732
Db        346 R 346
```

```
RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT:  DORNER, F.
;   APPLICANT:  SCHEIFLINGER, F.
;   APPLICANT:  FALKNER, F. G.
;   TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES:  52
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Foley & Lardner
;     STREET:  1800 Diagonal Road, Suite 500
;     CITY:  Alexandria
;     STATE:  VA
;     COUNTRY:  USA
;     ZIP:  22313-0299
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/232,463
;     FILING DATE:
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US/07/935,313
;     FILING DATE:
;     APPLICATION NUMBER:  EP 91 114 300.6
;     FILING DATE:  26-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  BENT, Stephen A.
;     REGISTRATION NUMBER:  29,768
;     REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (703)836-9300
;     TELEFAX:  (703)683-4109
;     TELEX:  899149
; INFORMATION FOR SEQ ID NO:  14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  7218 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   IMMEDIATE SOURCE:
;     CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             3.8%;  Score 43.8;  DB 2;  Length 7218;
  Best Local Similarity   14.8%;  Pred. No. 0.26;
  Matches   84;  Conservative  208;  Mismatches  275;  Indels    0;  Gaps    0;

Qy        180 ATTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCC 239
Db       1633 ATTCATTTATAGCATAGAAAAAAACAAAATGAAATTCTACTATATTTTTACATACATATA 1574

Qy        240 ACTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTC 299
Db       1573 TTCTAAATATGAAAGTGGTGATTGTGACTAGCGTAGCATCGCTTCTAGACGCATCTATTC 1514

Qy        300 CGTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCC 359
Db       1513 AGTTTCAAAAAACGGCATGTAGGCATCACTGTAATTACCTATCTATGCAAGTAGTTAAAG 1454

Qy        360 TATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGG 419
Db       1453 AGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1394

Qy        420 AATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAAC 479
Db       1393 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1334

Qy        480 CGCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAA 539
Db       1333 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1274

Qy        540 AGTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGT 599
Db       1273 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1214

Qy        600 AAAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTAT 659
Db       1213 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1154

Qy        660 TGATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGA 719
Db       1153 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1094

Qy        720 TGGACCGGAGATGAAAAAAATAAAGAG 746
Db       1093 RRRRRRRRRRRRRRRRRRRRRRRRRRR 1067
```

```
RESULT 7
US-09-902-540-1853
; Sequence 1853, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1853
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-1853
```

```
  Query Match             3.6%;  Score 42.2;  DB 3;  Length 456;
  Best Local Similarity   52.6%;  Pred. No. 0.24;
  Matches   92;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy        755 CAGACAATATAAAATTGCTCGGTTATCAATCTTTTCCTGTTTTAAAAGAGTATATGCAGA 814
              | || |||||    || || ||  ||   |    | |      |    | |      |
Db        137 CCGAAAATATTCGCTTCCTTGGCAATGTCTTGGATGCGTCGCTTCCGGCGCTATACCGAG 196

Qy        815 GCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTC 874
               |||  | ||| |  | ||  |  | || ||||| || || || |  ||  | ||| | |
Db        197 ACGCTCGCGCGCTCATCTTCACGCCCGAGGAGGATTTCGGCATCACCCCGCTGGAATCCC 256

Qy        875 AAGCTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGT 929
              | ||    ||    || || || ||||  |||||||| || ||  | || || ||
Db        257 AGGCGGCGGGCCGGCCCGTCATCGCCTACGGGAAGGGCGGCGCGCTCGAGACGGT 311


RESULT 8
US-09-902-540-746/c
; Sequence 746, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 746
;   LENGTH: 5752
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-746

  Query Match             3.6%;  Score 42.2;  DB 3;  Length 5752;
  Best Local Similarity   52.6%;  Pred. No. 0.62;
  Matches   92;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy        755 CAGACAATATAAAATTGCTCGGTTATCAATCTTTTCCTGTTTTAAAAGAGTATATGCAGA 814
              | || |||||    || || ||  ||   |    | |      |    | |      |
Db       5580 CCGAAAATATTCGCTTCCTTGGCAATGTCTTGGATGCGTCGCTTCCGGCGCTATACCGAG 5521

Qy        815 GCGCCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTC 874
               |||  | ||| |  | ||  |  | || ||||| || || || |  ||  | ||| | |
Db       5520 ACGCTCGCGCGCTCATCTTCACGCCCGAGGAGGATTTCGGCATCACCCCGCTGGAATCCC 5461

Qy        875 AAGCTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGT 929
              | ||    ||    || || || ||||  |||||||| || ||  | || || ||
Db       5460 AGGCGGCGGGCCGGCCCGTCATCGCCTACGGGAAGGGCGGCGCGCTCGAGACGGT 5406


RESULT 9
US-09-949-016-203910
; Sequence 203910, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 203910
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-203910

  Query Match             3.5%;  Score 41;  DB 3;  Length 601;
  Best Local Similarity   52.8%;  Pred. No. 0.55;
  Matches   86;  Conservative    1;  Mismatches   76;  Indels    0;  Gaps    0;

Qy        510 TTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGGCTTCAGTTATATA 569
              |||    |  ||:   | |   ||| | || || | | |   ||  || || | | || ||
Db        290 TTCAGTAATTAKTCTGAGATTAATAAGAATAGAAAAAATTAGTTTGGTTTAACTAATTTA 349

Qy        570 TCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAGACTATTATTTCAC 629
              |  |  |||  ||||  ||| |||  |||   |   | | ||  ||  | || |  ||
Db        350 TATGTGTGTGTATGTATATACTTTGTAAGGCTAGCTTAACAATGAAATCAATAAAGTCCT 409

Qy        630 AGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTC 672
               ||| ||   | || | | |||   |  | | || || || ||
Db        410 TGCAACCTAAAAGGAAACATACTGCCTAAGTAATTTTTTTATC 452


RESULT 10
US-09-949-016-12928
; Sequence 12928, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12928
;   LENGTH: 194537
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(194537)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12928

  Query Match             3.5%;  Score 41;  DB 3;  Length 194537;
  Best Local Similarity   52.0%;  Pred. No. 4.8;
  Matches   92;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy        110 TATATTCGGTGGTTGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATG 169
              | |||||       |||||  | ||  |       || |    ||||| |    || | ||
Db     126792 TTTATTCAAGTTAAGATTTAGTGACATACATGTTAATATCCAATTTTTTTTTCAATTGTG 126851

Qy        170 CGACTACCACATTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTAC 229
              |   ||| | |   |  | |||| |   | |||| ||| |   | |||| ||
Db     126852 TGGAGGCCAAAATATGTTTCACTTTAACCTCAACTAACAAAATCACCAAGAAGTAAGATG 126911

Qy        230 CACTAATGCCACTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTA 286
              | | | || |||    | | |||| || |||  ||    ||| |||||  |||| |
Db     126912 TAATGAAGCAACTCTTGACTTAACAGCTAGATAAATATATTGCAAATATATTCATAA 126968


RESULT 11
US-09-949-016-12740
```

```
; Sequence 12740, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12740
;   LENGTH: 201529
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(201529)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12740

  Query Match             3.5%;  Score 41;  DB 3;  Length 201529;
  Best Local Similarity   52.0%;  Pred. No. 4.9;
  Matches   92;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy        110 TATATTCGGTGGTTGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATG 169
              | |||||          |||||  | ||  |      || |     |||||  |     || | ||
Db     133792 TTTATTCAAGTTAAGATTTAGTGACATACATGTTAATATCCAATTTTTTTTTCAATTGTG 133851

Qy        170 CGACTACCACATTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTAC 229
              |    ||| | |   |   | |||| |   | |||| ||| |   | |||| ||
Db     133852 TGGAGGCCAAAATATGTTTCACTTTAACCTCAACTAACAAAATCACCAAGAAGTAAGATG 133911

Qy        230 CACTAATGCCACTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTA 286
              | | | || |||    | | |||| || |||  ||    ||| |||||  |||| |
Db     133912 TAATGAAGCAACTCTTGACTTAACAGCTAGATAAATATATTGCAAATATATTCATAA 133968


RESULT 12
US-09-949-016-12805/c
; Sequence 12805, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12805
;   LENGTH: 421491
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(421491)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12805

  Query Match             3.5%;  Score 40;  DB 3;  Length 421491;
  Best Local Similarity   46.2%;  Pred. No. 12;
  Matches  133;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        498 TTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGGC 557
              || | || ||||   |||| |          ||||    | ||  ||  | |    | |
Db     205361 TTATTTAACTAAAGTTCAAAACTCACTCAAAAGAAGCCTAGAAAGATTGACATCACATGT 205302

Qy        558 TTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAGA 617
              | ||  |  |  |||   |    |  |||| |      ||||||   | ||||||||
Db     205301 TGCAAATTGAATTCCTTGTAATCAAAAGGATGACAAATTAGTAAATTCTTAAAAGCAAAC 205242

Qy        618 CTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAAGC 677
                | | ||  | ||        | | |  |  || ||||  |||| ||   | | ||
Db     205241 AGGGCTTCTAACTACTTCTTTACCAGGAACTGAGTAATGTATCAATCTAATAACCAATGC 205182

Qy        678 CTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAAAA 737
              |||| | | | |     |||| ||    | |  |||   |       | | | |||  |
Db     205181 ATTTACTTATAGGAATACAAAGGAAAAGATTGCCATTTCAGCATTTTAGAACAGGAAGCA 205122

Qy        738 AATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTATCAATC 785
              |   ||    ||      || |||||| ||      ||||   | ||
Db     205121 GAACTAGCAACAGTTATTTGAAAATATACAAAAAACTGGTTTGAAGTC 205074


RESULT 13
US-09-949-016-14060/c
; Sequence 14060, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14060
;   LENGTH: 421494
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(421494)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14060

  Query Match             3.5%;  Score 40;  DB 3;  Length 421494;
  Best Local Similarity   46.2%;  Pred. No. 12;
  Matches  133;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        498 TTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGGC 557
              || | || ||||   |||| |          ||||    | ||  ||  | |    | |
Db     205361 TTATTTAACTAAAGTTCAAAACTCACTCAAAAGAAGCCTAGAAAGATTGACATCACATGT 205302

Qy        558 TTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAGA 617
              | ||  |  |  |||   |    |  |||| |      ||||||   | ||||||||
Db     205301 TGCAAATTGAATTCCTTGTAATCAAAAGGATGACAAATTAGTAAATTCTTAAAAGCAAAC 205242

Qy        618 CTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAAGC 677
                | | ||  | ||        | | |  |  || ||||  |||| ||   | | ||
Db     205241 AGGGCTTCTAACTACTTCTTTACCAGGAACTGAGTAATGTATCAATCTAATAACCAATGC 205182

Qy        678 CTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGACCGGAGATGAAAAA 737
              |||| | | | |     |||| ||    | |  |||   |       | | | |||  |
Db     205181 ATTTACTTATAGGAATACAAAGGAAAAGATTGCCATTTCAGCATTTTAGAACAGGAAGCA 205122
```

```
Qy        738 AATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTATCAATC 785
              |   ||    ||       || |||||| ||      ||||   | ||
Db     205121 GAACTAGCAACAGTTATTTGAAAATATACAAAAAACTGGTTTGAAGTC 205074


RESULT 14
US-09-949-016-17503/c
; Sequence 17503, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17503
;   LENGTH: 38119
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17503

  Query Match             3.4%;  Score 39.8;  DB 3;  Length 38119;
  Best Local Similarity   52.8%;  Pred. No. 5.4;
  Matches   86;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        510 TTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGGCTTCAGTTATATA 569
              |||    |  ||    | |   ||| | || || | | |   ||  || || | | || ||
Db      34219 TTCAGTAATTAGTCTGAGATTAATAAGAATAGAAAAAATTAGTTTGGTTTAACTAATTTA 34160

Qy        570 TCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAGACTATTATTTCAC 629
              |  |  |||  |||| ||| |||  |||    |   | | ||  ||  | || |  ||
Db      34159 TATGTGTGTGTATGTATATACTTTGTAAGGCTAGCTTAACAATGAAATCAATAAAGTCCT 34100

Qy        630 AGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTC 672
              ||| ||   | || | | |||   |  | | || || || ||
Db      34099 TGCAACCTAAAAGGAAACATACTGCCTAAGTAATTTTTTTATC 34057


RESULT 15
US-09-722-377-9/c
; Sequence 9, Application US/09722377
; Patent No. 6791007
; GENERAL INFORMATION:
;  APPLICANT: Schulze-Lefert, Paul MJ
;  APPLICANT:  Panstruga, Ralph
;  APPLICANT:  Buschges, Rainer
;  TITLE OF INVENTION: Polynucleotide and its use for modulating a defence
;  TITLE OF INVENTION:  response in plants
;  FILE REFERENCE: 620-58
;  CURRENT APPLICATION NUMBER: US/09/722,377
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER: PCT/GB97/02046
;  PRIOR FILING DATE: 1997-07-29
;  PRIOR APPLICATION NUMBER: GB 9615879.5
;  PRIOR FILING DATE: 1996-07-29
;  PRIOR APPLICATION NUMBER: GB 9622626.1
;  PRIOR FILING DATE: 1996-10-30
;  PRIOR APPLICATION NUMBER: GB 9704789.8
;  PRIOR FILING DATE: 1997-03-07
;  NUMBER OF SEQ ID NOS: 79
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 7175
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-722-377-9

  Query Match             3.4%;  Score 39.4;  DB 3;  Length 7175;
  Best Local Similarity   49.8%;  Pred. No. 3.7;
  Matches  100;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        950 AACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTG 1009
              ||||||  ||||| || | |||   |||   || |    |  |      || ||| | |
Db       2882 AACCGATGGGCATGTTATCCAACATACAATCTAATCACACCGTTATATTAGATGTAATTA 2823

Qy       1010 AATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAAT 1069
              ||||| | || ||   ||||  || |     ||    |   ||||  |  ||    |||
Db       2822 AATTTTACAACAAGATATCACATTATTACATTCTAATTAAAAGAATATGTGTTTTAAACC 2763

Qy       1070 TTTCTCGATCAAGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTT 1129
              ||      |||| |||| |   | |||   || |  ||   |   |    |  | | | |
Db       2762 GTTAAAACTCAATATTTCATACATAATAGCAGTATGTTTGAACATATGTCTTAAGTGTGT 2703

Qy       1130 TCAAAACAGAACAGATTATTA 1150
              |  | | |   || || | ||
Db       2702 TTCACAAAATTCATATAACTA 2682


Search completed: May 12, 2006, 03:51:38
Job time : 166.379 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:31:18 ; Search time 741.451 Seconds
(without alignments)
12915.142 Million cell updates/sec

Title: US-09-423-093B-3_COPY_7310_8467
Perfect score: 1158
Sequence: 1 atggaaataaatgaaataaa.....aacagattattaaacgttaa 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1158	100.0	12441	3	US-09-423-093A-3	Sequence 3, Appli
c 2	53	4.6	852	4	US-09-925-065A-67989	Sequence 67989, A
c 3	53	4.6	1814	4	US-09-925-065A-727044	Sequence 727044,
c 4	53	4.6	1814	4	US-09-925-065A-727045	Sequence 727045,
c 5	53	4.6	14096	8	US-10-719-993-6916	Sequence 6916, Ap
6	53	4.6	193691	8	US-10-719-993-6768	Sequence 6768, Ap
c 7	53	4.6	318760	8	US-10-719-993-6765	Sequence 6765, Ap
c 8	44.6	3.9	598	5	US-10-027-632-243685	Sequence 243685,
c 9	44.6	3.9	598	6	US-10-027-632-243685	Sequence 243685,
10	43.4	3.7	32767	9	US-10-948-947A-98	Sequence 98, Appl
11	42.4	3.7	1756	5	US-10-235-994-27	Sequence 27, Appl
12	42.4	3.7	1766	9	US-10-450-763-27184	Sequence 27184, A
13	42.4	3.7	2406	9	US-10-450-763-27186	Sequence 27186, A
14	42.4	3.7	2491	9	US-10-498-788-75	Sequence 75, Appl
15	42.2	3.6	2000	3	US-09-938-842A-4142	Sequence 4142, Ap
16	42.2	3.6	2000	3	US-09-938-842A-4142	Sequence 4142, Ap
17	42.2	3.6	11155	6	US-10-311-455-577	Sequence 577, App
18	41.6	3.6	3673778	6	US-10-312-841-1	Sequence 1, Appli
19	41.2	3.6	6000	6	US-10-311-455-2277	Sequence 2277, Ap
c 20	41.2	3.6	3673778	6	US-10-312-841-1	Sequence 1, Appli
21	41	3.5	1889	3	US-09-981-353-83	Sequence 83, Appl
c 22	41	3.5	139326	7	US-10-322-281-134	Sequence 134, App
c 23	40.8	3.5	408	9	US-10-275-323A-13	Sequence 13, Appl
24	40.2	3.5	6136	7	US-10-257-166-124	Sequence 124, App
c 25	40.2	3.5	19659	6	US-10-311-455-740	Sequence 740, App
c 26	40	3.5	6113	6	US-10-311-455-775	Sequence 775, App
c 27	40	3.5	6127	6	US-10-311-455-566	Sequence 566, App
c 28	39.8	3.4	391	7	US-10-424-599-36734	Sequence 36734, A
c 29	39.8	3.4	15165	10	US-11-097-143-28507	Sequence 28507, A
30	39.6	3.4	1603	4	US-09-925-065A-65905	Sequence 65905, A
31	39.6	3.4	1992	7	US-10-282-122A-16177	Sequence 16177, A
c 32	39.4	3.4	520	4	US-09-925-065A-528376	Sequence 528376,
c 33	39.4	3.4	530	4	US-09-925-065A-308763	Sequence 308763,
34	39.4	3.4	603	5	US-10-027-632-212961	Sequence 212961,
35	39.4	3.4	603	6	US-10-027-632-212961	Sequence 212961,
36	39.4	3.4	2083	7	US-10-424-599-99586	Sequence 99586, A
c 37	39.4	3.4	4031	10	US-11-097-143-37933	Sequence 37933, A
38	39.4	3.4	6203	5	US-10-239-676-184	Sequence 184, App
39	39.4	3.4	6203	6	US-10-240-453-274	Sequence 274, App
c 40	39.4	3.4	8923	7	US-10-221-714A-61	Sequence 61, Appl
c 41	39	3.4	627	4	US-09-925-065A-744572	Sequence 744572,
c 42	39	3.4	627	4	US-09-925-065A-819131	Sequence 819131,
c 43	39	3.4	627	4	US-09-925-065A-819132	Sequence 819132,
44	39	3.4	648	5	US-10-027-632-281751	Sequence 281751,
45	39	3.4	648	5	US-10-027-632-281753	Sequence 281753,

ALIGNMENTS

RESULT 1
US-09-423-093A-3
; Sequence 3, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12441
; TYPE: DNA
; ORGANISM: Salmonella enterica
US-09-423-093A-3

Query Match 100.0%; Score 1158; DB 3; Length 12441;
Best Local Similarity 100.0%; Pred. No. 2e-270;
Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7310 ATGGAAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGC 7369

Qy         61 TCCGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7370 TCCGAACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTG 7429

Qy        121 GTTGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7430 GTTGATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACA 7489

Qy        181 TTTATTCAAAATTTACCTAAAGCTAAAAAAATTTTACCAGAAATATTTACCACTAATGCCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7490 TTTATTCAAAATTTACCTAAAGCTAAAAAAATTTTACCAGAAATATTTACCACTAATGCCA 7549

Qy        241 CTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7550 CTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCC 7609
```

```
Qy        301 GTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCT 7669

Qy        361 ATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 ATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGA 7729

Qy        421 ATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 ATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACC 7789

Qy        481 GCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAA 7849

Qy        541 GTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 GTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTA 7909

Qy        601 AAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 AAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATT 7969

Qy        661 GATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGAT 8029

Qy        721 GGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTAT 8089

Qy        781 CAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCG 8149

Qy        841 GAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCC 8209

Qy        901 TTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGC 8269

Qy        961 ATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 ATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAA 8329

Qy       1021 AATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 AATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCA 8389

Qy       1081 AGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGATTTGAACAAGAATTTAAGAACTTTGTTAATGAAAAGTGGAATCTTTTCAAAACAGAA 8449

Qy       1141 CAGATTATTAAACGTTAA 1158
              ||||||||||||||||||
Db       8450 CAGATTATTAAACGTTAA 8467
```

```
RESULT 2
US-09-925-065A-67989/c
; Sequence 67989, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 67989
;   LENGTH: 852
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-67989

  Query Match             4.6%;  Score 53;  DB 4;  Length 852;
  Best Local Similarity   49.8%;  Pred. No. 0.028;
  Matches  134;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        818 CCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAG 877
              || |||  | | || || |   | | |   |||| | |||| |  | ||||||||
Db        524 CCTGGGACTGTTTTCTTCCCAGTGGACTGTTTGTTGGTAGAATAACCCCCAAAGCTCAAA 465

Qy        878 CTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCAC 937
              |   | | |  | | |  | | || |  | |  |  || ||||  | | ||  |
Db        464 GCTAAAATGCATCATTAGTCCTCGTCGGCAGTTGAGCTGATTTGGAAAAGCTGCACCTTC 405

Qy        938 TAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATG 997
              |  | |||  ||| |||  ||  | |        | | ||    |  |||  || | |
Db        404 CTGCAGAAGATCAACTGACCTGCTATCCCACTGCACATCTGAGGTATATTTCAGTGAAGG 345

Qy        998 AAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAA 1057
              || |   |||  ||| | |  | | || |||| | | |  |   |  || |||
Db        344 CAGGTAGCTGTGCTTTTCAGAGCAGAGAAGCAGTTTTAAGAGCAGAAAGGTAGAGGAAAT 285

Qy       1058 ATGCAGAAAAATTTTCTCGATCAAGATTT 1086
              |  | || ||   ||| |||  ||||||
Db        284 CTAGAAAAGAACCGTCTTGATACAGATTT 256


RESULT 3
US-09-925-065A-727044/c
; Sequence 727044, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 727044
;   LENGTH: 1814
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-727044

  Query Match             4.6%;  Score 53;  DB 4;  Length 1814;
```

```
  Best Local Similarity   49.8%;  Pred. No. 0.041;
  Matches  134;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        818 CCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAG 877
              || |||  | | || || |    | |  |    |||| | |||| |   | ||||||||
Db        524 CCTGGGACTGTTTTCTTCCCAGTGGACTGTTTGTTGGTAGAATAACCCCCAAAGCTCAAA 465

Qy        878 CTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCAC 937
              |    | |  |  |  |   | | || |   | |  |  || ||||   | | ||  |
Db        464 GCTAAAATGCATCATTAGTCCTCGTCGGCAGTTGAGCTGATTTGGAAAAGCTGCACCTTC 405

Qy        938 TAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATG 997
              |  | |||  ||| ||| ||  |  |       | | ||    |  |||  || | |
Db        404 CTGCAGAAGATCAACTGACCTGCTATCCCACTGCACATCTGAGGTATATTTCAGTGAAGG 345

Qy        998 AAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAA 1057
              || |    |||  ||| | |  |  | | || |||| | | |  |    |  || |||
Db        344 CAGGTAGCTGTGCTTTTCAGAGCAGAGAAGCAGTTTTAAGAGCAGAAAGGTAGAGGAAAT 285

Qy       1058 ATGCAGAAAAATTTTCTCGATCAAGATTT 1086
              |  | || ||   ||| |||  ||||||
Db        284 CTAGAAAAGAACCGTCTTGATACAGATTT 256


RESULT 4
US-09-925-065A-727045/c
; Sequence 727045, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 727045
;   LENGTH: 1814
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-727045

  Query Match             4.6%;  Score 53;  DB 4;  Length 1814;
  Best Local Similarity   49.8%;  Pred. No. 0.041;
  Matches  134;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        818 CCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAG 877
              || |||  | | || || |    | |  |    |||| | |||| |   | ||||||||
Db        524 CCTGGGACTGTTTTCTTCCCAGTGGACTGTTTGTTGGTAGAATAACCCCCAAAGCTCAAA 465

Qy        878 CTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCAC 937
              |    | |  |  |  |   | | || |   | |  |  || ||||   | | ||  |
Db        464 GCTAAAATGCATCATTAGTCCTCGTCGGCAGTTGAGCTGATTTGGAAAAGCTGCACCTTC 405

Qy        938 TAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATG 997
              |  | |||  ||| ||| ||  |  |       | | ||    |  |||  || | |
Db        404 CTGCAGAAGATCAACTGACCTGCTATCCCACTGCACATCTGAGGTATATTTCAGTGAAGG 345

Qy        998 AAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAA 1057
              || |    |||  ||| | |  |  | | || |||| | | |  |    |  || |||
Db        344 CAGGTAGCTGTGCTTTTCAGAGCAGAGAAGCAGTTTTAAGAGCAGAAAGGTAGAGGAAAT 285

Qy       1058 ATGCAGAAAAATTTTCTCGATCAAGATTT 1086
              |  | || ||   ||| |||  ||||||
Db        284 CTAGAAAAGAACCGTCTTGATACAGATTT 256


RESULT 5
US-10-719-993-6916/c
; Sequence 6916, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6916
;   LENGTH: 14096
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-6916

  Query Match             4.6%;  Score 53;  DB 8;  Length 14096;
  Best Local Similarity   49.8%;  Pred. No. 0.11;
  Matches  134;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        818 CCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAG 877
              || |||  | | || || |    | |  |    |||| | |||| |   | ||||||||
Db       7724 CCTGGGACTGTTTTCTTCCCAGTGGACTGTTTGTTGGTAGAATAACCCCCAAAGCTCAAA 7665

Qy        878 CTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCAC 937
              |    | |  |  |  |   | | || |   | |  |  || ||||   | | ||  |
Db       7664 GCTAAAATGCATCATTAGTCCTCGTCGGCAGTTGAGCTGATTTGGAAAAGCTGCACCTTC 7605

Qy        938 TAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATG 997
              |  | |||  ||| ||| ||  |  |       | | ||    |  |||  || | |
Db       7604 CTGCAGAAGATCAACTGACCTGCTATCCCACTGCACATCTGAGGTATATTTCAGTGAAGG 7545

Qy        998 AAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAA 1057
              || |    |||  ||| | |  |  | | || |||| | | |  |    |  || |||
Db       7544 CAGGTAGCTGTGCTTTTCAGAGCAGAGAAGCAGTTTTAAGAGCAGAAAGGTAGAGGAAAT 7485

Qy       1058 ATGCAGAAAAATTTTCTCGATCAAGATTT 1086
              |  | || ||   ||| |||  ||||||
Db       7484 CTAGAAAAGAACCGTCTTGATACAGATTT 7456


RESULT 6
US-10-719-993-6768
; Sequence 6768, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6768
;   LENGTH: 193691
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(193691)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-3
US-10-719-993-6768

  Query Match             4.6%;  Score 53;  DB 8;  Length 193691;
```

```
  Best Local Similarity   49.8%;  Pred. No. 0.4;
  Matches  134;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        818 CCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAG 877
              || |||  | | || ||  |    | |  |    |||| | |||| |   | ||||||||
Db     173730 CCTGGGACTGTTTTCTTCCCAGTGGACTGTTTGTTGGTAGAATAACCCCCAAAGCTCAAA 173789

Qy        878 CTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCAC 937
              |    | |  |  | |   | | || |   | |  |  || ||||  | | ||  |
Db     173790 GCTAAAATGCATCATTAGTCCTCGTCGGCAGTTGAGCTGATTTGGAAAAGCTGCACCTTC 173849

Qy        938 TAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATG 997
              |  | |||  ||| |||  ||  |  |       | | ||    |  |||  || | |
Db     173850 CTGCAGAAGATCAACTGACCTGCTATCCCACTGCACATCTGAGGTATATTTCAGTGAAGG 173909

Qy        998 AAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAA 1057
              || |    |||  ||| | |  |  | | || |||| | | |  |    |  || |||
Db     173910 CAGGTAGCTGTGCTTTTCAGAGCAGAGAAGCAGTTTTAAGAGCAGAAAGGTAGAGGAAAT 173969

Qy       1058 ATGCAGAAAAATTTTCTCGATCAAGATTT 1086
              |  | || ||   ||| |||  ||||||
Db     173970 CTAGAAAAGAACCGTCTTGATACAGATTT 173998


RESULT 7
US-10-719-993-6765/c
; Sequence 6765, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6765
;   LENGTH: 318760
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(318760)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6765

  Query Match             4.6%;  Score 53;  DB 8;  Length 318760;
  Best Local Similarity   49.8%;  Pred. No. 0.51;
  Matches  134;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        818 CCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAG 877
              || |||  | | || ||  |    | |  |    |||| | |||| |   | ||||||||
Db     191840 CCTGGGACTGTTTTCTTCCCAGTGGACTGTTTGTTGGTAGAATAACCCCCAAAGCTCAAA 191781

Qy        878 CTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCAC 937
              |    | |  |  | |   | | || |   | |  |  || ||||  | | ||  |
Db     191780 GCTAAAATGCATCATTAGTCCTCGTCGGCAGTTGAGCTGATTTGGAAAAGCTGCACCTTC 191721

Qy        938 TAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATG 997
              |  | |||  ||| |||  ||  |  |       | | ||    |  |||  || | |
Db     191720 CTGCAGAAGATCAACTGACCTGCTATCCCACTGCACATCTGAGGTATATTTCAGTGAAGG 191661

Qy        998 AAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAA 1057
              || |    |||  ||| | |  |  | | || |||| | | |  |    |  || |||
Db     191660 CAGGTAGCTGTGCTTTTCAGAGCAGAGAAGCAGTTTTAAGAGCAGAAAGGTAGAGGAAAT 191601

Qy       1058 ATGCAGAAAAATTTTCTCGATCAAGATTT 1086
              |  | || ||   ||| |||  ||||||
Db     191600 CTAGAAAAGAACCGTCTTGATACAGATTT 191572


RESULT 8
US-10-027-632-243685/c
; Sequence 243685, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 243685
;   LENGTH: 598
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-243685

  Query Match             3.9%;  Score 44.6;  DB 5;  Length 598;
  Best Local Similarity   58.8%;  Pred. No. 2.6;
  Matches   77;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy        536 AAAAAGTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTG 595
              ||||| ||| || | ||||    | | ||||| |||   ||| | |  | || ||| |
Db        400 AAAAACTATTCAAAGGTGAAAGATAATTTATAAATCGTGCTGGAAAAATTGAAAATATGT 341

Qy        596 AAGTAAAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAAC 655
              | | |||||    |||   ||||  |||||| ||| | | |   |   |   ||   |
Db        340 AGGCAAAAATAATAAATTTAGACACTTATTTTACATCTTTCACAAAAATTAACTCATAGT 281

Qy        656 GTATTGATCTT 666
              | ||||||| |
Db        280 GGATTGATCAT 270


RESULT 9
US-10-027-632-243685/c
; Sequence 243685, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
```

```
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 243685
;   LENGTH: 598
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-243685

  Query Match             3.9%;  Score 44.6;  DB 6;  Length 598;
  Best Local Similarity   58.8%;  Pred. No. 2.6;
  Matches   77;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy        536 AAAAAGTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTG 595
              ||||| ||| || | ||||    | | ||||| |||   ||| | |  | || ||| |
Db        400 AAAAACTATTCAAAGGTGAAAGATAATTTATAAATCGTGCTGGAAAAATTGAAAATATGT 341

Qy        596 AAGTAAAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAAC 655
              | | |||||    |||   ||||  |||||| ||| | | |   |   |   ||   |
Db        340 AGGCAAAAATAATAAATTTAGACACTTATTTTACATCTTTCACAAAAATTAACTCATAGT 281

Qy        656 GTATTGATCTT 666
              | ||||||| |
Db        280 GGATTGATCAT 270


RESULT 10
US-10-948-947A-98
; Sequence 98, Application US/10948947A
; Publication No. US20050130924A1
; GENERAL INFORMATION:
;  APPLICANT: Monia, Brett P.
;  APPLICANT:  Freier, Susan M.
;  APPLICANT:  Manoharan, Muthiah
;  APPLICANT:  Gaarde, William A.
;  APPLICANT:  Griffey, Richard H.
;  APPLICANT:  Swayze, Eric E.
;  APPLICANT:  Bennett, C. Frank
;  TITLE OF INVENTION: ANTISENSE INHIBITION VIA RNASE H-INDEPENDENT REDUCTION IN mRNA
;  FILE REFERENCE: ISPH-0871
;  CURRENT APPLICATION NUMBER: US/10/948,947A
;  CURRENT FILING DATE:  2004-09-24
;  PRIOR APPLICATION NUMBER: 60/392,020
;  PRIOR FILING DATE: 2002-06-26
;  PRIOR APPLICATION NUMBER: 10/461,163
;  PRIOR FILING DATE: 2003-06-13
;  NUMBER OF SEQ ID NOS: 115
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 98
;   LENGTH: 32767
;   TYPE: DNA
;   ORGANISM: M. musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 7748-7847
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 11218-11318
;   OTHER INFORMATION: n = A,T,C or G
US-10-948-947A-98

  Query Match             3.7%;  Score 43.4;  DB 9;  Length 32767;
  Best Local Similarity   47.6%;  Pred. No. 36;
  Matches  166;  Conservative    0;  Mismatches  176;  Indels    7;  Gaps    1;

Qy        713 TTGGTGATGGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGC 772
              |||| |||  |  |  |    || |   ||  || ||| |   |||  | | |    |||
Db      25176 TTGGGGATTTAGAGAGGTCCCAATAGGGAAGTAGAAAGACATTAGAAGACACATTTCTGC 25235

Qy        773 TCGGTTATCAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTT 832
                |      |||||| | |   ||||| | |    | ||| || |          || |
Db      25236 CTGAGCTGAAATCTTATGCCTGTTTAACATATCTAAAGCACAGGGAGGAAATTCTTTTCA 25295

Qy        833 TTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTG 892
              || | ||  |    |||||          ||    |    | | | | ||  | |  | |
Db      25296 TTCCTGCCTATAGTGACTTCCTGCCCCTAGAATTTAGGGATTAGGTTTATGCTGTCTCCT 25355

Qy        893 TTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAAC 952
              || |||  ||| | |  ||  |   |   |    ||   || |  |  |  |    ||
Db      25356 TTGTTGTATTTCAGTATAGTTAGAGGTGGCATTGGGTGGACCTAGGAACTTGATTTGAGT 25415

Qy        953 CGACTGGCATTTTCTTC-------AAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTT 1005
                |  |||||| |||       || ||||    | ||  ||  |||   | |||  ||
Db      25416 TTCCAAGCATTTGATTCCCAATTTAATGAACCATCTCTTTATTAGTTGAGAGCAGCCTTT 25475

Qy       1006 AGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAA 1054
              |||| || ||||   | | | |  |  |||  | || ||||||  ||||
Db      25476 AGTGCATATGAACTTATTCCCTTGTCATTTGGAACTGAGGCTTTCAGAA 25524


RESULT 11
US-10-235-994-27
; Sequence 27, Application US/10235994
; Publication No. US20030101002A1
; GENERAL INFORMATION:
;  APPLICANT: Bartha, Gabor
;  APPLICANT:  Walker, Michael
;  TITLE OF INVENTION: METHODS FOR ANALYZING GENE EXPRESSION PATTERNS
;  FILE REFERENCE: ICYTP012
;  CURRENT APPLICATION NUMBER: US/10/235,994
;  CURRENT FILING DATE:  2002-09-04
;  PRIOR APPLICATION NUMBER: US/10/003,608
;  PRIOR FILING DATE: 2001-11-01
;  PRIOR APPLICATION NUMBER: 60/245,081
;  PRIOR FILING DATE: 2000-11-01
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 1756
;   TYPE: DNA
;   ORGANISM: Human
US-10-235-994-27

  Query Match             3.7%;  Score 42.4;  DB 5;  Length 1756;
  Best Local Similarity   49.1%;  Pred. No. 15;
  Matches  112;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        883 GGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGT 942
              || ||| || ||  ||    ||    |  | || |  | |||     ||     ||| |
Db        138 GGAACCACTCTTGGTGGGAATGTTTTGATTTGGCCAATGGAAGGTAGTCATTGGCTAAAT 197

Qy        943 GTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCT 1002
              ||  ||  |   |  |       || | ||  |  || ||| ||   ||  | |  |  ||
Db        198 GTTAAGATAATTATAGATGAGCTCATTAAAAAGGAGCATAATGTGACTGTCCTAGTTGCC 257

Qy       1003 GTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCA 1062
                | ||| | ||   |  |   || | | |     ||||| | | | |  || | ||
Db        258 TCTGGTGCACTTTTCATCACACCAACCTCTAACCCATCTCTGACATTTGAAATATATAAG 317

Qy       1063 GAAAAATTTTCTCGATCAAGATTTGAACAAGAATTTAAGAACTTTGTT 1110
              |     |||      |  |||| | |||  || | ||||| |||| |||
Db        318 GTGCCCTTTGGCAAAGAAAGAATAGAAGGAGTAATTAAGGACTTCGTT 365


RESULT 12
US-10-450-763-27184
; Sequence 27184, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
;  FILE REFERENCE: 790CIP3/US
;  CURRENT APPLICATION NUMBER: US/10/450,763
;  CURRENT FILING DATE:  2003-06-11
;  PRIOR APPLICATION NUMBER: PCT/US01/08631
```

```
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 09/540,217
;  PRIOR FILING DATE: 2000-03-31
;  PRIOR APPLICATION NUMBER: 09/649,167
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 60736
;  SOFTWARE: Custom
; SEQ ID NO 27184
;   LENGTH: 1766
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SIMILAR
;   LOCATION: (64)..(1644)
;   OTHER INFORMATION: 100% homologous to Homo sapiens UDP
;   OTHER INFORMATION: glucuronosyltransferase,accession number AJ006054,Smith-Waterman
;   OTHER INFORMATION: Score=2798.
US-10-450-763-27184

  Query Match             3.7%;  Score 42.4;  DB 9;  Length 1766;
  Best Local Similarity   49.1%;  Pred. No. 15;
  Matches  112;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        883 GGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGT 942
              || ||| || ||  ||     ||     | | || |  | |||     ||     ||| |
Db        109 GGAACCACTCTTGGTGGGAATGTTTTGATTTGGCCAATGGAAGGTAGTCATTGGCTAAAT 168

Qy        943 GTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCT 1002
              ||  ||  |   | |       || | ||  |  || ||| ||   || | |   | ||
Db        169 GTTAAGATAATTATAGATGAGCTCATTAAAAAGGAGCATAATGTGACTGTCCTAGTTGCC 228

Qy       1003 GTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCA 1062
              | ||| | ||   |  |   || | | |     ||||| | | | |  || | ||
Db        229 TCTGGTGCACTTTTCATCACACCAACCTCTAACCCATCTCTGACATTTGAAATATATAAG 288

Qy       1063 GAAAAATTTTCTCGATCAAGATTTGAACAAGAATTTAAGAACTTTGTT 1110
              |     |||      |  |||| | |||  || | ||||| |||| |||
Db        289 GTGCCCTTTGGCAAAGAAAGAATAGAAGGAGTAATTAAGGACTTCGTT 336


RESULT 13
US-10-450-763-27186
; Sequence 27186, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
;  FILE REFERENCE: 790CIP3/US
;  CURRENT APPLICATION NUMBER: US/10/450,763
;  CURRENT FILING DATE:  2003-06-11
;  PRIOR APPLICATION NUMBER: PCT/US01/08631
;  PRIOR FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 09/540,217
;  PRIOR FILING DATE: 2000-03-31
;  PRIOR APPLICATION NUMBER: 09/649,167
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 60736
;  SOFTWARE: Custom
; SEQ ID NO 27186
;   LENGTH: 2406
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SIMILAR
;   LOCATION: (1117)..(1644)
;   OTHER INFORMATION: 84% homologous to Rattus norvegicus UDP-
;   OTHER INFORMATION: glucuronosyltransferase 2A1 precursor,accession number X57565,Smi
;   OTHER INFORMATION: -Waterman Score=800.
US-10-450-763-27186

  Query Match             3.7%;  Score 42.4;  DB 9;  Length 2406;
  Best Local Similarity   49.1%;  Pred. No. 18;
  Matches  112;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        883 GGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGT 942
              || ||| || ||  ||     ||     | | || |  | |||     ||     ||| |
Db       1135 GGAACCACTCTTGGTGGGAATGTTTTGATTTGGCCAATGGAAGGTAGTCATTGGCTAAAT 1194

Qy        943 GTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCT 1002
              ||  ||  |   | |       || | ||  |  || ||| ||   || | |   | ||
Db       1195 GTTAAGATAATTATAGATGAGCTCATTAAAAAGGAGCATAATGTGACTGTCCTAGTTGCC 1254

Qy       1003 GTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCA 1062
              | ||| | ||   |  |   || | | |     ||||| | | | |  || | ||
Db       1255 TCTGGTGCACTTTTCATCACACCAACCTCTAACCCATCTCTGACATTTGAAATATATAGG 1314

Qy       1063 GAAAAATTTTCTCGATCAAGATTTGAACAAGAATTTAAGAACTTTGTT 1110
              |     |||      |  |||| | |||  || | ||||| |||| |||
Db       1315 GTGCCCTTTGGCAAAGAAAGAATAGAAGGAGTAATTAAGGACTTCGTT 1362


RESULT 14
US-10-498-788-75
; Sequence 75, Application US/10498788
; Publication No. US20050118594A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  Chawla, Narinder K.
;  APPLICANT:  Lee, Soo Yeun
;  APPLICANT:  Ring, Huijun Z.
;  APPLICANT:  Lee, Ernestine A.
;  APPLICANT:  Forsythe, Ian J.
;  APPLICANT:  Khare, Reena
;  APPLICANT:  Tran, Uyen K.
;  APPLICANT:  Kable, Amy E.
;  APPLICANT:  Richardson, Thomas W.
;  APPLICANT:  Emerling, Brooke M.
;  APPLICANT:  Lindquist, Erika A.
;  APPLICANT:  Baughn, Mariah R.
;  APPLICANT:  Hafalia, April J. A.
;  APPLICANT:  Jin, Pei
;  APPLICANT:  Swarnakar, Anita
;  APPLICANT:  Li, Joana X.
;  APPLICANT:  Marquis, Joseph P.
;  APPLICANT:  Lee, Sally
;  APPLICANT:  Gorvad, Ann E.
;  APPLICANT:  Sprague, William W.
;  APPLICANT:  Becha, Shanya D.
;  APPLICANT:  Elliott, Vicki S.
;  TITLE OF INVENTION: ENZYMES
;  FILE REFERENCE: PF-1312 PCT
;  CURRENT APPLICATION NUMBER: US/10/498,788
;  CURRENT FILING DATE:  2004-06-14
;  PRIOR APPLICATION NUMBER: US 60/340,357
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/342,962
;  PRIOR FILING DATE: 2001-12-20
;  PRIOR APPLICATION NUMBER: US 60/343,558
;  PRIOR FILING DATE: 2001-12-21
;  PRIOR APPLICATION NUMBER: US 60/351,107
;  PRIOR FILING DATE: 2002-01-22
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: PERL Program
; SEQ ID NO 75
;   LENGTH: 2491
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 7508332CB1
US-10-498-788-75

  Query Match             3.7%;  Score 42.4;  DB 9;  Length 2491;
  Best Local Similarity   49.1%;  Pred. No. 18;
  Matches  112;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;
```

```
Qy        883 GGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGT 942
              || ||| || ||  ||    ||    |  | || |  | |||     ||     |||  |
Db        138 GGAACCACTCTTGGTGGGAATGTTTTGATTTGGCCAATGGAAGGTAGTCATTGGCTAAAT 197

Qy        943 GTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCT 1002
              ||  ||  |   |  |       || | ||  |  || ||| ||   || | |  |  ||
Db        198 GTTAAGATAATTATAGATGAGCTCATTAAAAAGGAGCATAATGTGACTGTCCTAGTTGCC 257

Qy       1003 GTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCA 1062
                | ||| | ||   |  |   || | | |    ||||| | | | |  || | ||
Db        258 TCTGGTGCACTTTTCATCACACCAACCTCTAACCCATCTCTGACATTTGAAATATATAAG 317

Qy       1063 GAAAAATTTTCTCGATCAAGATTTGAACAAGAATTTAAGAACTTTGTT 1110
              |     |||     |  |||| | |||  || | ||||| |||| |||
Db        318 GTGCCCTTTGGCAAAGAAAGAATAGAAGGAGTAATTAAGGACTTCGTT 365
```

RESULT 15
US-09-938-842A-4142
; Sequence 4142, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4142
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4142

Query Match 3.6%; Score 42.2; DB 3; Length 2000;
Best Local Similarity 51.3%; Pred. No. 18;
Matches 98; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```
Qy        499 TTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTATACAGACGTGAGGCT 558
              ||| |||  ||   | |||| ||||  ||    ||||| ||| |  | |  | |     |
Db        622 TTTGTAGTAAAAATTGAATATGTCGCTTGTTTGGTTAATAAACTTCAAATGCTTATTTTT 681

Qy        559 TCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAAATGAAAAGCAAGAC 618
                || || |    |    | |  |||||  | || || | | ||||  |||   |  |
Db        682 GTTTTTCTAAAAAATCTACAAAAATTGGATGCTGTTAAAATGATAAATCTAAATGCAATC 741

Qy        619 TATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATCTTATTGTCGAAGCC 678
              |     |  | | |||   || |  |||   ||||| |   |  |  || ||  || |
Db        742 TCAATGTAGAAAACATATTGTCTAATACTACACAAACTTGCGGTCCAAATGGTAAAACCA 801

Qy        679 TTTAGTAAAAT 689
              | || || | |
Db        802 TATACTATACT 812
```

Search completed: May 12, 2006, 05:34:31
Job time : 747.451 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:52:36 ; Search time 590 Seconds
(without alignments)
8004.956 Million cell updates/sec

Title: US-09-423-093B-3_COPY_7310_8467
Perfect score: 1158
Sequence: 1 atggaaataaatgaaataaa.....aacagattattaaacgttaa 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/2/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/2/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	165.8	14.3	1457619	17	US-11-098-686-8739	Sequence 8739, Ap
	2	163.2	14.1	1104	17	US-11-098-686-9413	Sequence 9413, Ap
c	3	59.6	5.1	200	17	US-11-098-686-4581	Sequence 4581, Ap
c	4	53	4.6	852	7	US-09-925-065A-67989	Sequence 67989, A
c	5	53	4.6	852	11	US-10-301-480-169228	Sequence 169228,
c	6	53	4.6	852	12	US-10-301-480-782637	Sequence 782637,
c	7	53	4.6	1814	7	US-09-925-065A-727044	Sequence 727044,
c	8	53	4.6	1814	7	US-09-925-065A-727045	Sequence 727045,
c	9	52.4	4.5	200	17	US-11-098-686-4578	Sequence 4578, Ap
	10	42.4	3.7	1715	10	US-10-750-185-24635	Sequence 24635, A
	11	42.4	3.7	1715	10	US-10-750-623-24635	Sequence 24635, A
c	12	40.4	3.5	976	12	US-10-301-480-548093	Sequence 548093,
c	13	40.4	3.5	976	12	US-10-301-480-1161502	Sequence 1161502,
	14	39.8	3.4	491	11	US-10-301-480-20345	Sequence 20345, A
	15	39.8	3.4	491	12	US-10-301-480-633754	Sequence 633754,
	16	39.6	3.4	1603	7	US-09-925-065A-65905	Sequence 65905, A
	17	39.6	3.4	1603	11	US-10-301-480-167144	Sequence 167144,
	18	39.6	3.4	1603	12	US-10-301-480-780553	Sequence 780553,
c	19	39.4	3.4	520	7	US-09-925-065A-528376	Sequence 528376,
c	20	39.4	3.4	530	7	US-09-925-065A-308763	Sequence 308763,
c	21	39.4	3.4	565	11	US-10-301-480-61849	Sequence 61849, A
c	22	39.4	3.4	565	12	US-10-301-480-675258	Sequence 675258,
c	23	39	3.4	627	7	US-09-925-065A-744572	Sequence 744572,
c	24	39	3.4	627	7	US-09-925-065A-819131	Sequence 819131,
c	25	39	3.4	627	7	US-09-925-065A-819132	Sequence 819132,
	26	38.8	3.4	538	12	US-10-301-480-568470	Sequence 568470,
	27	38.8	3.4	538	12	US-10-301-480-1181879	Sequence 1181879,
c	28	38.8	3.4	577	12	US-10-301-480-480903	Sequence 480903,
c	29	38.8	3.4	577	12	US-10-301-480-1094312	Sequence 1094312,
c	30	38.8	3.4	608	7	US-09-925-065A-415189	Sequence 415189,
c	31	38.6	3.3	604	7	US-09-925-065A-217691	Sequence 217691,
c	32	38.6	3.3	604	7	US-09-925-065A-217692	Sequence 217692,
c	33	38.6	3.3	611	12	US-10-301-480-303920	Sequence 303920,
c	34	38.6	3.3	611	12	US-10-301-480-303921	Sequence 303921,
c	35	38.6	3.3	611	12	US-10-301-480-917329	Sequence 917329,
c	36	38.6	3.3	611	12	US-10-301-480-917330	Sequence 917330,
c	37	38.2	3.3	200	17	US-11-098-686-4580	Sequence 4580, Ap
	38	38.2	3.3	4891	13	US-10-204-639-136	Sequence 136, App
c	39	37.8	3.3	403	11	US-10-301-480-43775	Sequence 43775, A
c	40	37.8	3.3	403	12	US-10-301-480-657184	Sequence 657184,
c	41	37.8	3.3	414	11	US-10-301-480-43776	Sequence 43776, A
c	42	37.8	3.3	414	11	US-10-301-480-43777	Sequence 43777, A
c	43	37.8	3.3	414	12	US-10-301-480-657185	Sequence 657185,
c	44	37.8	3.3	414	12	US-10-301-480-657186	Sequence 657186,
c	45	37.6	3.2	648	7	US-09-925-065A-509163	Sequence 509163,

ALIGNMENTS

RESULT 1
US-11-098-686-8739/c
; Sequence 8739, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8739
; LENGTH: 1457619
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-8739

Query Match 14.3%; Score 165.8; DB 17; Length 1457619;
Best Local Similarity 49.7%; Pred. No. 7.6e-26;
Matches 552; Conservative 0; Mismatches 532; Indels 27; Gaps 4;

```
Qy          5 AAATAAATGAAATAAAAATATCTCTCGTTCATGAGTGGTTATTAAGTTATGCAGGCTCCG 64
              ||| |  | | ||    ||  || | |||||| | ||| |  || |     |||    |
Db     628760 AAAGAGGTAATATGCGGATTGCTATTGTTCATTATTGGCTTATAGGAATGCGAGGAGGAG 628701

Qy         65 AACAGGTATCATCTGCCATCCTGCATGTTTTTCCTGAAGCGAAGTTATATTCGGTGGTTG 124
                | | ||     |  |    | | |||  ||| ||||| |  | | | |      ||
Db     628700 AGAAAGTTATTGAGTCTCTTTGTCGTCTTTACCCTCAAGCGGATATTTTTACTCACTGTG 628641

Qy        125 ATTTTCTAACGGATGAACAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATTTA 184
                  |    | |||| |||    |      ||       ||   |     ||| |||||||
```

```
Db     628640 ---TAGATAAGGATCAACTTTCACCTTTATTAAAAAATAAGACTATTTATACTACATTTA 628584

Qy        185 TTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGG 244

Db     628583 TTAATAGTCTTCCTTTTGCAACTCGGTACTATAAGTATTATCTTCCGCTTATGCCTTTAG 628524

Qy        245 CTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTG 304

Db     628523 CTCTAGAACAGCTTGATCTTCGGGGGTATGATTTAATAATTTCCAGCGAATCCGGCCCTG 628464

Qy        305 CAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCATTCTCCTATTC 364

Db     628463 CAAAAGGACTTATTGTTCATTCTTCTGCTCGACATATATGTTATTGTCATACTCCCATGA 628404

Qy        365 GATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAATAAAGGAATTA 424

Db     628403 GGTATTTATGGGACTGTTATCACACTTATCTTGAGAATAAAGGTATAGTAACACGAACTA 628344

Qy        425 AAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTCTCGAACCGCAA 484

Db     628343 TCATTCG---------TCCTATTTTTCATTACTTACGGATGTGGGATATAAGTTCTGCGG 628293

Qy        485 ATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAAAAAGTAT 544

Db     628292 TACACCCTGATATTATTCTTGCAAATTCTCATACAGTAGCACAACGTGTGCAACGTTGCT 628233

Qy        545 ACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAAGTAAAAA 604

Db     628232 GGGGACGTTCAGCTTCTGTTATTTATCCACCTGTTGAGCTTGACCTATTTAAACCCACAA 628173

Qy        605 ATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATTGATC 664

Db     628172 GAGATGTTGGTGATTATTACCTATGTTTTGGTCAGCTTGTAGAGTATAAGAGGGTAGATT 628113

Qy        665 TTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGATGGAC 724

Db     628112 TGGCTATTAAAGTTGCATCGGAAA---CAGGTCGTAAATTAATTGTTGCTGGTGATGGAG 628056

Qy        725 CGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTATCAAT 784

Db     628055 ATGCATATACAAAACTTAAACAACTGGCAGGTCCAAGTGTTATTTTTATAGGTAAAATTT 627996

Qy        785 CTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTTTGCAGCGGAAG 844

Db     627995 CTATGGACGAAAGGGCTAAATTACTAGCTGGTTGTCGTGCACTACTCTTCCCTGGGGAAG 627936

Qy        845 AGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGTTATTGCCTTTG 904

Db     627935 AAGATTTTGGTATAGTTCCTATAGAAGCTATGGCATCAGGAAGACCTGTCATTGCTTATA 627876

Qy        905 GGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACCGACTGGCATTT 964

Db     627875 GCAAAGGAGGGGCATCAGAAAC-----------AGTTTCTGATGGTATAACAGGTTTTT 627828

Qy        965 TCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATTTGAAAAAAATG 1024

Db     627827 TCTTTCATGAACAAACACCAGAAGCTTTATTAAAGGCTATTGAGTATTTTGAAAAACATG 627768

Qy       1025 CATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTCTCGATCAAGAT 1084

Db     627767 AACATACCATTAATACTCAGCAGCTTGTAAAACATGCCTCTAAATTTAGTGAGACTCAAT 627708

Qy       1085 TTGAACAAGAATTTAAGAACTTTGTTAATGA 1115

Db     627707 TTCATCAAGCGTTTTCAAAAGTTGTTGATAA 627677
```

RESULT 2
US-11-098-686-9413
; Sequence 9413, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9413
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-9413

Query Match 14.1%; Score 163.2; DB 17; Length 1104;
Best Local Similarity 51.0%; Pred. No. 5.3e-26;
Matches 480; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

```
Qy        174 TACCACATTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACT 233

Db        156 TACTACATTTATTAATAGTCTTCCTTTTGCAACTCGGTACTATAAGTATTATCTTCCGCT 215

Qy        234 AATGCCACTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGC 293

Db        216 TATGCCTTTAGCTCTAGAACAGCTTGATCTTCGGGGGTATGATTTAATAATTTCCAGCGA 275

Qy        294 CCATTCCGTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCA 353

Db        276 ATCCGGCCCTGCAAAAGGACTTATTGTTCATTCTTCTGCTCGACATATATGTTATTGTCA 335

Qy        354 TTCTCCTATTCGATATGCGTGGGATTTACAGCATCAGTACCTTAATGAGTCTAACCTGAA 413

Db        336 TACTCCCATGAGGTATTTATGGGACTGTTATCACACTTATCTTGAGAATAAAGGTATAGT 395

Qy        414 TAAAGGAATTAAAGGTTGGTTAGCAAAATGGCTTCTTCACAAAATACGAATTTGGGATTC 473

Db        396 AACACGAACTATCATTCG---------TCCTATTTTTCATTACTTACGGATGTGGGATAT 446

Qy        474 TCGAACCGCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAAT 533

Db        447 AAGTTCTGCGGTACACCCTGATATTATTCTTGCAAATTCTCATACAGTAGCACAACGTGT 506

Qy        534 TAAAAAAGTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTT 593

Db        507 GCAACGTTGCTGGGGACGTTCAGCTTCTGTTATTTATCCACCTGTTGAGCTTGACCTATT 566

Qy        594 TGAAGTAAAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAA 653

Db        567 TAAACCCACAAGAGATGTTGGTGATTATTACCTATGTTTTGGTCAGCTTGTAGAGTATAA 626

Qy        654 ACGTATTGATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTAT 713

Db        627 GAGGGTAGATTTGGCTATTAAAGTTGCATCGGAAA---CAGGTCGTAAATTAATTGTTGC 683

Qy        714 TGGTGATGGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCT 773

Db        684 TGGTGATGGAGATGCATATACAAAACTTAAACAACTGGCAGGTCCAAGTGTTATTTTTAT 743

Qy        774 CGGTTATCAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCCAGGGCGTTTGTTTT 833

Db        744 AGGTAAAATTTCTATGGACGAAAGGGCTAAATTACTAGCTGGTTGTCGTGCACTACTCTT 803

Qy        834 TGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTTGCGGTACCCCTGT 893

Db        804 CCCTGGGGAAGAAGATTTTGGTATAGTTCCTATAGAAGCTATGGCATCAGGAAGACCTGT 863

Qy        894 TATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCACTAGGTGTAGAGGAACC 953

Db        864 CATTGCTTATAGCAAAGGAGGGGCATCAGAAACAGTTTCT-----------GATGGTAT 911
```

```
Qy        954 GACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATGAAGCTGTTAGTGAATT 1013
              || ||  |||||||  | |||||  |      |   ||||     | ||| ||    | ||
Db        912 AACAGGTTTTTTCTTTCATGAACAAACACCAGAAGCTTTATTAAAGGCTATTGAGTATTT 971

Qy       1014 TGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAAATGCAGAAAAATTTTC 1073
              |||||| ||| |        |||   |||||      |  |||| ||||    ||||||
Db        972 TGAAAAACATGAACATACCATTAATACTCAGCAGCTTGTAAAACATGCCTCTAAATTTAG 1031

Qy       1074 TCGATCAAGATTTGAACAAGAATTTAAGAACTTTGTTAATGA 1115
              |    |  |||| | ||||  |||   ||  ||||| || |
Db       1032 TGAGACTCAATTTCATCAAGCGTTTTCAAAAGTTGTTGATAA 1073
```

RESULT 3
US-11-098-686-4581/c
; Sequence 4581, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4581
; LENGTH: 200
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
US-11-098-686-4581

Query Match 5.1%; Score 59.6; DB 17; Length 200;
Best Local Similarity 56.7%; Pred. No. 0.0019;
Matches 110; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
Qy        174 TACCACATTTATTCAAAATTTACCTAAAGCTAAAAAATTTTACCAGAAATATTTACCACT 233
              ||| |||||||||| | | | | |||   || |      |  ||  || | ||| | || ||
Db        194 TACTACATTTATTAATAGTCTTCCTTTTGCAACTCGGTACTATAAGTATTATCTTCCGCT 135

Qy        234 AATGCCACTGGCTATTGAACAACTTGATTTATCAGATGCTAATATCATCATTAGTAGCGC 293
              |||||  | ||| | ||||| |||||| |    |    | || | || |||   ||||
Db        134 TATGCCTTTAGCTCTAGAACAGCTTGATCTTCGGGGGTATGATTTAATAATTTCCAGCGA 75

Qy        294 CCATTCCGTTGCAAAAGGTGTTATTTCCGGACCAGATCAGCTTCACATTAGCTATGTTCA 353
              |  |||||||||  |||||       |   |   |  || ||  | |||  |||
Db         74 ATCCGGCCCTGCAAAAGGACTTATTGTTCATTCTTCTGCTCGACATATATGTTATTGTCA 15

Qy        354 TTCTCCTATTCGAT 367
              | |||| ||  | |
Db         14 TACTCCCATGAGGT 1
```

RESULT 4
US-09-925-065A-67989/c
; Sequence 67989, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 67989
; LENGTH: 852
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-67989

Query Match 4.6%; Score 53; DB 7; Length 852;
Best Local Similarity 49.8%; Pred. No. 0.075;
Matches 134; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
Qy        818 CCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAG 877
              || |||  | | || || |    | |  |    |||| | |||| |   | ||||||||
Db        524 CCTGGGACTGTTTTCTTCCCAGTGGACTGTTTGTTGGTAGAATAACCCCCAAAGCTCAAA 465

Qy        878 CTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCAC 937
              |    | | |  | |   | | || |   | |  |  || ||||   | | ||  |
Db        464 GCTAAAATGCATCATTAGTCCTCGTCGGCAGTTGAGCTGATTTGGAAAAGCTGCACCTTC 405

Qy        938 TAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATG 997
              |  | |||  ||| |||  ||  | |        | | ||    |  |||  || | |
Db        404 CTGCAGAAGATCAACTGACCTGCTATCCCACTGCACATCTGAGGTATATTTCAGTGAAGG 345

Qy        998 AAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAA 1057
              || |    |||  ||| | |  |  | | || |||| | | |  |    |  || |||
Db        344 CAGGTAGCTGTGCTTTTCAGAGCAGAGAAGCAGTTTTAAGAGCAGAAAGGTAGAGGAAAT 285

Qy       1058 ATGCAGAAAAATTTTCTCGATCAAGATTT 1086
              |  | || ||   ||| |||  ||||||
Db        284 CTAGAAAAGAACCGTCTTGATACAGATTT 256
```

RESULT 5
US-10-301-480-169228/c
; Sequence 169228, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 169228
; LENGTH: 852
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-169228

Query Match 4.6%; Score 53; DB 11; Length 852;
Best Local Similarity 49.8%; Pred. No. 0.075;
Matches 134; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
Qy        818 CCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAG 877
              || |||  | | || || |    | |  |    |||| | |||| |   | ||||||||
Db        524 CCTGGGACTGTTTTCTTCCCAGTGGACTGTTTGTTGGTAGAATAACCCCCAAAGCTCAAA 465

Qy        878 CTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCAC 937
              |    | |  |  | |   | | || |   | |  |  || ||||   | | ||  |
Db        464 GCTAAAATGCATCATTAGTCCTCGTCGGCAGTTGAGCTGATTTGGAAAAGCTGCACCTTC 405
```

```
Qy        938 TAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATG 997
              |  | |||  ||| |||  ||  |  |       | | ||    |  |||  || | |
Db        404 CTGCAGAAGATCAACTGACCTGCTATCCCACTGCACATCTGAGGTATATTTCAGTGAAGG 345

Qy        998 AAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAA 1057
              || |    |||  ||| | |  | | | || |||| | | |  |    |  || |||
Db        344 CAGGTAGCTGTGCTTTTCAGAGCAGAGAAGCAGTTTTAAGAGCAGAAAGGTAGAGGAAAT 285

Qy       1058 ATGCAGAAAAATTTTCTCGATCAAGATTT 1086
              |  | || ||   ||| |||  ||||||
Db        284 CTAGAAAAGAACCGTCTTGATACAGATTT 256


RESULT 6
US-10-301-480-782637/c
; Sequence 782637, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 782637
;   LENGTH: 852
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-782637

  Query Match             4.6%;  Score 53;  DB 12;  Length 852;
  Best Local Similarity   49.8%;  Pred. No. 0.075;
  Matches  134;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        818 CCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAG 877
              || |||  | | || ||  |    | |  |    |||| | |||| |   | |||||||||
Db        524 CCTGGGACTGTTTTCTTCCCAGTGGACTGTTTGTTGGTAGAATAACCCCCAAAGCTCAAA 465

Qy        878 CTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCAC 937
              |    | |  |  |  |  | | || |   | |  |  || ||||  | | ||  |
Db        464 GCTAAAATGCATCATTAGTCCTCGTCGGCAGTTGAGCTGATTTGGAAAAGCTGCACCTTC 405

Qy        938 TAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATG 997
              |  | |||  ||| |||  ||  |  |       | | ||    |  |||  || | |
Db        404 CTGCAGAAGATCAACTGACCTGCTATCCCACTGCACATCTGAGGTATATTTCAGTGAAGG 345

Qy        998 AAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAA 1057
              || |    |||  ||| | |  | | | || |||| | | |  |    |  || |||
Db        344 CAGGTAGCTGTGCTTTTCAGAGCAGAGAAGCAGTTTTAAGAGCAGAAAGGTAGAGGAAAT 285

Qy       1058 ATGCAGAAAAATTTTCTCGATCAAGATTT 1086
              |  | || ||   ||| |||  ||||||
Db        284 CTAGAAAAGAACCGTCTTGATACAGATTT 256


RESULT 7
US-09-925-065A-727044/c
; Sequence 727044, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 727044
;   LENGTH: 1814
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-727044

  Query Match             4.6%;  Score 53;  DB 7;  Length 1814;
  Best Local Similarity   49.8%;  Pred. No. 0.09;
  Matches  134;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        818 CCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAG 877
              || |||  | | || ||  |    | |  |    |||| | |||| |   | |||||||||
Db        524 CCTGGGACTGTTTTCTTCCCAGTGGACTGTTTGTTGGTAGAATAACCCCCAAAGCTCAAA 465

Qy        878 CTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCAC 937
              |    | |  |  |  |  | | || |   | |  |  || ||||  | | ||  |
Db        464 GCTAAAATGCATCATTAGTCCTCGTCGGCAGTTGAGCTGATTTGGAAAAGCTGCACCTTC 405

Qy        938 TAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATG 997
              |  | |||  ||| |||  ||  |  |       | | ||    |  |||  || | |
Db        404 CTGCAGAAGATCAACTGACCTGCTATCCCACTGCACATCTGAGGTATATTTCAGTGAAGG 345

Qy        998 AAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAA 1057
              || |    |||  ||| | |  | | | || |||| | | |  |    |  || |||
Db        344 CAGGTAGCTGTGCTTTTCAGAGCAGAGAAGCAGTTTTAAGAGCAGAAAGGTAGAGGAAAT 285

Qy       1058 ATGCAGAAAAATTTTCTCGATCAAGATTT 1086
              |  | || ||   ||| |||  ||||||
Db        284 CTAGAAAAGAACCGTCTTGATACAGATTT 256


RESULT 8
US-09-925-065A-727045/c
; Sequence 727045, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 727045
;   LENGTH: 1814
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-727045

  Query Match             4.6%;  Score 53;  DB 7;  Length 1814;
  Best Local Similarity   49.8%;  Pred. No. 0.09;
```

```
  Matches  134;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        818 CCAGGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAG 877
              || |||  | | || || |    | |   |    |||| | |||| |   | ||||||||
Db        524 CCTGGGACTGTTTTCTTCCCAGTGGACTGTTTGTTGGTAGAATAACCCCCAAAGCTCAAA 465

Qy        878 CTTGCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTTCGCCCAC 937
              |   | | |  |  |  | | || |   | |  |  || ||||   | | ||  |
Db        464 GCTAAAATGCATCATTAGTCCTCGTCGGCAGTTGAGCTGATTTGGAAAAGCTGCACCTTC 405

Qy        938 TAGGTGTAGAGGAACCGACTGGCATTTTCTTCAAGGAACAGAATATTGCTTCTTTGCATG 997
                | | |||  ||| |||  ||  |  |       | | ||    |  |||  || | |
Db        404 CTGCAGAAGATCAACTGACCTGCTATCCCACTGCACATCTGAGGTATATTTCAGTGAAGG 345

Qy        998 AAGCTGTTAGTGAATTTGAAAAAAATGCATCATTTTTTACATCTCAGGCTTGTAGAAAAA 1057
              || |    |||  ||| | |  |  | | || |||| | | |  |    |  || |||
Db        344 CAGGTAGCTGTGCTTTTCAGAGCAGAGAAGCAGTTTTAAGAGCAGAAAGGTAGAGGAAAT 285

Qy       1058 ATGCAGAAAAATTTTCTCGATCAAGATTT 1086
                |  | || ||   ||| |||  ||||||
Db        284 CTAGAAAAGAACCGTCTTGATACAGATTT 256


RESULT 9
US-11-098-686-4578/c
; Sequence 4578, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
;  APPLICANT: Kapur, Vivek and Gebhart, Connie J.
;  TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
;  TITLE OF INVENTION:  FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
;  FILE REFERENCE: 09531-128001
;  CURRENT APPLICATION NUMBER: US/11/098,686
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: PCT/US03/31318
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: US 60/416,395
;  PRIOR FILING DATE: 2002-10-04
;  NUMBER OF SEQ ID NOS: 11433
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4578
;   LENGTH: 200
;   TYPE: DNA
;   ORGANISM: Lawsonia intracellularis
US-11-098-686-4578

  Query Match             4.5%;  Score 52.4;  DB 17;  Length 200;
  Best Local Similarity   67.3%;  Pred. No. 0.073;
  Matches   74;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy        821 GGGCGTTTGTTTTTGCAGCGGAAGAGGACTTTGGAATAATACCTGTCGAAGCTCAAGCTT 880
              | ||  |  | ||  | | |||||| || ||||| ||| | ||| | ||||||   || |
Db        159 GTGCACTACTCTTCCCTGGGGAAGAAGATTTTGGTATAGTTCCTATAGAAGCTATGGCAT 100

Qy        881 GCGGTACCCCTGTTATTGCCTTTGGGAAGGGTGGGGCCTTAGAAACCGTT 930
              || |  ||||| ||||| | | | || || ||||| | |||||| |||
Db         99 CAGGAAGACCTGTCATTGCTTATAGCAAAGGAGGGGCATCAGAAACAGTT 50


RESULT 10
US-10-750-185-24635
; Sequence 24635, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 24635
;   LENGTH: 1715
;   TYPE: DNA
;   ORGANISM: Bovine  19866880360692
US-10-750-185-24635

  Query Match             3.7%;  Score 42.4;  DB 10;  Length 1715;
  Best Local Similarity   50.5%;  Pred. No. 19;
  Matches  103;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        142 CAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATTTATTCAAAATTTACCTAAA 201
              ||||| | | |||||     | | ||    | |  |  ||  | |   ||  ||  || |
Db       1330 CAAAGCATAGATTTTGCTTTGTATTACCAAAGTTACTTATCAAGTTTTAAAATATTTAGA 1389

Qy        202 GCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCTATTGAACAACTTGAT 261
                  | |||||||| ||| || || | ||| || |    |   | | | ||| ||  || |
Db       1390 TTATAGAAATTTTAACAGTAACATCTTCCAGTATTTAGATATGTTCTAGAAGAAAGTGTT 1449

Qy        262 TTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCAAAAGGTGTTATTTCC 321
              || |    |   ||  |  |  ||  |      || |   |||   |   | | ||||||
Db       1450 TTCTTTCTTCAAAACCTATTTTTTTTTTCAAGTCAGTGTATTGTTCATCATTTCATTTCC 1509

Qy        322 GGACCAGATCAGCTTCACATTAGC 345
                  |   | |||| | | || ||
Db       1510 TATCTTTAAAAGCTACCCCTTGGC 1533


RESULT 11
US-10-750-623-24635
; Sequence 24635, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-1
;  CURRENT APPLICATION NUMBER: US/10/750,623
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 24635
;   LENGTH: 1715
;   TYPE: DNA
;   ORGANISM: Bovine  19866880360692
US-10-750-623-24635

  Query Match             3.7%;  Score 42.4;  DB 10;  Length 1715;
  Best Local Similarity   50.5%;  Pred. No. 19;
  Matches  103;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        142 CAAAGAAGACATTTTCTGGGGAAATATGCGACTACCACATTTATTCAAAATTTACCTAAA 201
              ||||| | | |||||     | | ||    | |  |  ||  | |   ||  ||  || |
Db       1330 CAAAGCATAGATTTTGCTTTGTATTACCAAAGTTACTTATCAAGTTTTAAAATATTTAGA 1389

Qy        202 GCTAAAAAATTTTACCAGAAATATTTACCACTAATGCCACTGGCTATTGAACAACTTGAT 261
                  | |||||||| ||| || || | ||| || |    |   | | | ||| ||  || |
Db       1390 TTATAGAAATTTTAACAGTAACATCTTCCAGTATTTAGATATGTTCTAGAAGAAAGTGTT 1449

Qy        262 TTATCAGATGCTAATATCATCATTAGTAGCGCCCATTCCGTTGCAAAAGGTGTTATTTCC 321
              || |    |   ||  |  |  ||  |      || |   |||   |   | | ||||||
```

```
Db       1450 TTCTTTCTTCAAAACCTATTTTTTTTTTCAAGTCAGTGTATTGTTCATCATTTCATTTCC 1509

Qy        322 GGACCAGATCAGCTTCACATTAGC 345
                |   |  |||| | | || ||
Db       1510 TATCTTTAAAAGCTACCCCTTGGC 1533


RESULT 12
US-10-301-480-548093/c
; Sequence 548093, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 548093
;   LENGTH: 976
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-548093

  Query Match             3.5%;  Score 40.4;  DB 12;  Length 976;
  Best Local Similarity   45.2%;  Pred. No. 45;
  Matches  149;  Conservative    0;  Mismatches  181;  Indels    0;  Gaps    0;

Qy        478 ACCGCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAA 537
              |||    | | |||||   | |  ||  |   |  || |||   |   |  | |    |
Db        538 ACCCTTCAGGAGGTTGTAAAGTCCATTTCACCTAGTCTATACCACAGACTATGGAAACTA 479

Qy        538 AAAGTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAA 597
              |  |||| | |  |     |  |  |||||||    |||   | | ||  ||   | |
Db        478 TATATATATATATATATATATATATATATATATATAGTGCTGTCCCTCTAGAGAACCCTAAT 419

Qy        598 GTAAAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGT 657
              |  | |   || |   | |  ||     |  ||  |       |||| |  ||||||  ||
Db        418 ATGTATATATAATATATATAACATGCCCTGTCATTAAGATCAATGGGAAACTACAACAGT 359

Qy        658 ATTGATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGT 717
              ||  ||  |   | |  |     | ||   | ||      | |   ||    ||  |||
Db        358 ATATATATATATATATATATATATATATATATATATAATATATAAAATATATATATAATA 299

Qy        718 GATGGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGT 777
              |   |    | |  | | | |  | |   |  ||| | | ||||||  |||  |   |
Db        298 TAGATAATATATACTATATATAATAGTATATATACTATATATAATATATTATTTATGTAT 239

Qy        778 TATCAATCTTTTCCTGTTTTAAAAGAGTAT 807
              |||  |  || |  | | ||| |   ||||
Db        238 TATATACATTATATTATATTATATATGTAT 209


RESULT 13
US-10-301-480-1161502/c
; Sequence 1161502, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1161502
;   LENGTH: 976
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1161502

  Query Match             3.5%;  Score 40.4;  DB 12;  Length 976;
  Best Local Similarity   45.2%;  Pred. No. 45;
  Matches  149;  Conservative    0;  Mismatches  181;  Indels    0;  Gaps    0;

Qy        478 ACCGCAAATGGGGTTGATCATTTTATAGCTAATTCTCAATATATCGCGCGTAGAATTAAA 537
              |||    | | |||||   | |  ||  |   |  || |||   |   |  | |    |
Db        538 ACCCTTCAGGAGGTTGTAAAGTCCATTTCACCTAGTCTATACCACAGACTATGGAAACTA 479

Qy        538 AAAGTATACAGACGTGAGGCTTCAGTTATATATCCGCCTGTAGATGTGGATAATTTTGAA 597
              |  |||| | |  |     |  |  |||||||    |||   | | ||  ||   | |
Db        478 TATATATATATATATATATATATATATATATATATAGTGCTGTCCCTCTAGAGAACCCTAAT 419

Qy        598 GTAAAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGT 657
              |  | |   || |   | |  ||     |  ||  |       |||| |  ||||||  ||
Db        418 ATGTATATATAATATATATAACATGCCCTGTCATTAAGATCAATGGGAAACTACAACAGT 359

Qy        658 ATTGATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGT 717
              ||  ||  |   | |  |     | ||   | ||      | |   ||    ||  |||
Db        358 ATATATATATATATATATATATATATATATATATATAATATATAAAATATATATATAATA 299

Qy        718 GATGGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGT 777
              |   |    | |  | | | |  | |   |  ||| | | ||||||  |||  |   |
Db        298 TAGATAATATATACTATATATAATAGTATATATACTATATATAATATATTATTTATGTAT 239

Qy        778 TATCAATCTTTTCCTGTTTTAAAAGAGTAT 807
              |||  |  || |  | | ||| |   ||||
Db        238 TATATACATTATATTATATTATATATGTAT 209


RESULT 14
US-10-301-480-20345
; Sequence 20345, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20345
;   LENGTH: 491
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-20345

  Query Match             3.4%;  Score 39.8;  DB 11;  Length 491;
  Best Local Similarity   48.9%;  Pred. No. 52;
  Matches  107;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        601 AAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATT 660
              | |||| || |  |  |  | ||  |   |  || |  |       | | |   | |||
Db        103 AGAAATAAAGAATACAATAACTAAATTTAAAAATTCATTGGAAGCTTCAATAGTAGAATT 162

Qy        661 GATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGAT 720
              ||||      |  | |     | ||| |  |      | |||  ||| | | |||||     |
Db        163 GATCAAGCAGAAGGAAGAATCAGTGAGCTTAAAAATAAGCTATTTGAAATTATTCAATCT 222
```

```
Qy        721 GGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTAT 780
              | |  | | |  |||| |  |   |  ||| ||| ||  | |||| |||  |  |   |
Db        223 GAAGAGCAAAAAAAAATAGAATGAAARAAGTCTATAGGAATTATAGAATATCAGGTGAAC 282

Qy        781 CAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCC 819
               ||| ||| |  ||||   |||   | | | | |  |||
Db        283 AAATATTTACAGGTTTCTGAAGTAAAAAAGAAAAAAGCC 321


RESULT 15
US-10-301-480-633754
; Sequence 633754, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 633754
;   LENGTH: 491
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-633754

  Query Match             3.4%;  Score 39.8;  DB 12;  Length 491;
  Best Local Similarity   48.9%;  Pred. No. 52;
  Matches  107;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        601 AAAAATGAAAAGCAAGACTATTATTTCACAGCATCCCGTATGGTACCCTACAAACGTATT 660
              | |||| || |  |  |  | ||  |   |  || |  |        | | |   | |||
Db        103 AGAAATAAAGAATACAATAACTAAATTTAAAAATTCATTGGAAGCTTCAATAGTAGAATT 162

Qy        661 GATCTTATTGTCGAAGCCTTTAGTAAAATGCCGGAAAAGAAATTAGTAGTTATTGGTGAT 720
              ||||     |  | |    | ||| |  |     | |||  ||| | | |||||     |
Db        163 GATCAAGCAGAAGGAAGAATCAGTGAGCTTAAAAATAAGCTATTTGAAATTATTCAATCT 222

Qy        721 GGACCGGAGATGAAAAAAATAAAGAGCAAGGCTACAGACAATATAAAATTGCTCGGTTAT 780
              | |  | | |  |||| |  |   |  ||| ||| ||  | |||| |||  |  |   |
Db        223 GAAGAGCAAAAAAAAATAGAATGAAARAAGTCTATAGGAATTATAGAATATCAGGTGAAC 282

Qy        781 CAATCTTTTCCTGTTTTAAAAGAGTATATGCAGAGCGCC 819
               ||| ||| |  ||||   |||   | | | | |  |||
Db        283 AAATATTTACAGGTTTCTGAAGTAAAAAAGAAAAAAGCC 321


Search completed: May 12, 2006, 06:55:04
Job time : 593 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 08:31:04 ; Search time 4826.73 Seconds
(without alignments)
15792.702 Million cell updates/sec

Title: US-09-423-093B-3_COPY_1019_2359
Perfect score: 1341
Sequence: 1 atgaatcgtattattagaat.....tcaagagttcttatgagtaa 1341

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1341	100.0	12441	1	SERFBC2	X61917 S.enterica
	2	1341	100.0	12441	6	BD085556	BD085556 Nucleic a
	3	1341	100.0	12446	6	AX029263	AX029263 Sequence
	4	91	6.8	110000	1	CP000076_58	Continuation (59 o
c	5	64.4	4.8	254050	2	PFA929358	AL929358 Plasmodiu
c	6	56.8	4.2	266544	2	AC116956	AC116956 Dictyoste
	7	56.2	4.2	1522	2	AY701231	AY701231 Orconecte
	8	56.2	4.2	250078	2	AE014829	AE014829 Plasmodiu
	9	56	4.2	1141	6	AR579680	AR579680 Sequence
	10	56	4.2	1141	6	AX083744	AX083744 Sequence
c	11	55.6	4.1	250823	2	AE014821	AE014821 Plasmodiu
c	12	55	4.1	83313	14	AC143489	AC143489 Macaca mu
	13	54.4	4.1	146570	2	AC117076	AC117076 Dictyoste
	14	54.4	4.1	331039	2	AC116979	AC116979 Dictyoste
c	15	54	4.0	1141	6	AR579680	AR579680 Sequence
c	16	54	4.0	1141	6	AX083744	AX083744 Sequence
c	17	53	4.0	211380	14	CR388168	CR388168 Danio rer
	18	53	4.0	214325	5	BX571973	BX571973 Zebrafish
c	19	52.4	3.9	1522	2	AY701231	AY701231 Orconecte
	20	52.4	3.9	110000	14	AL954295_2	Continuation (3 of
	21	52.4	3.9	125031	5	BX248494	BX248494 Zebrafish
c	22	52.2	3.9	154986	14	AC141766	AC141766 Apis mell
c	23	52.2	3.9	253001	2	AE014834	AE014834 Plasmodiu
	24	51.8	3.9	172918	5	CR384100	CR384100 Zebrafish
c	25	51.6	3.8	112760	5	BX678772	BX678772 Zebrafish
	26	51	3.8	162047	5	BX784036	BX784036 Zebrafish
	27	50.6	3.8	19131	6	AX251474	AX251474 Sequence
	28	50.4	3.8	8923	6	AX251094	AX251094 Sequence
	29	50.4	3.8	160029	5	BX000984	BX000984 Zebrafish
	30	50.2	3.7	17234	6	AX458502	AX458502 Sequence
	31	50.2	3.7	110000	1	AE017263_4	Continuation (5 of
	32	50.2	3.7	310003	1	AE014016	AE014016 Buchnera
	33	49.8	3.7	15723	2	AY521251	AY521251 Aleurodic
	34	49.8	3.7	21537	6	AX346901	AX346901 Sequence
	35	49.8	3.7	209698	5	CR381585	CR381585 Zebrafish
c	36	49.8	3.7	348657	14	BX548252	BX548252 Danio rer
c	37	49.6	3.7	235457	14	CR450774	CR450774 Danio rer
	38	49.6	3.7	250029	2	AE014820	AE014820 Plasmodiu
	39	49.6	3.7	333321	2	AC116986	AC116986 Dictyoste
	40	49.4	3.7	7255	6	AX346180	AX346180 Sequence
	41	49.4	3.7	12138	6	AX346530	AX346530 Sequence
	42	49.4	3.7	12138	6	AX348420	AX348420 Sequence
	43	49.4	3.7	22977	8	AC106883	AC106883 Homo sapi
	44	49.2	3.7	8123	6	AX347129	AX347129 Sequence
c	45	49.2	3.7	169530	5	BX004756	BX004756 Zebrafish

ALIGNMENTS

RESULT 1
SERFBC2
LOCUS SERFBC2 12441 bp DNA linear BCT 09-SEP-2004
DEFINITION S.enterica rfbJ gene cluster.
ACCESSION X61917
VERSION X61917.1 GI:47004
KEYWORDS abe gene; abequose synthase; galactosyltransferase; manB gene; manC gene; mannose-1-phosphate guanyltransferase; phosphomannomutase; wbaL gene; wbaP gene; wbaQ gene; wbaR gene; wbaW gene; wbaZ gene; wzx gene; wzy gene.
SOURCE Salmonella enterica
ORGANISM Salmonella enterica
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 12441)
AUTHORS Brown,P.K., Romana,L.K. and Reeves,P.R.
TITLE Molecular analysis of the rfb gene cluster of Salmonella serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B
JOURNAL Mol. Microbiol. 6 (10), 1385-1394 (1992)
PUBMED 1379320
REFERENCE 2
AUTHORS Reeves,P.
TITLE Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale
JOURNAL Trends Genet. 9 (1), 17-22 (1993)
PUBMED 8434412
REFERENCE 3
AUTHORS Liu,D., Haase,A.M., Lindqvist,L., Lindberg,A.A. and Reeves,P.R.
TITLE Glycosyl transferases of O-antigen biosynthesis in Salmonella enterica: identification and characterization of transferase genes of groups B, C2, and E1
JOURNAL J. Bacteriol. 175 (11), 3408-3413 (1993)
PUBMED 7684736
REFERENCE 4 (bases 1 to 12441)
AUTHORS Brown,P.K.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1991) P.K. Brown, University of Sydney, Dept of Microbiology, Maze Crescent, Sydney N S W 2006, AUSTRALIA
FEATURES Location/Qualifiers
source 1. .12441

/organism="Salmonella enterica"
/mol_type="genomic DNA"
/strain="group C2, serovar muenchen M67"
/db_xref="taxon:28901"
gene 138. .1019
/gene="abe"
CDS 138. .1019
/gene="abe"
/codon_start=1
/transl_table=11
/product="abequose synthase"
/protein_id="CAA43918.1"
/db_xref="GI:47005"
/db_xref="GOA:Q00329"
/db_xref="UniProt/Swiss-Prot:Q00329"
/translation="MLDVNKKILMTGATSFVGTHLLHSLIKEGYSIIALKRPITEPTI
INTLIEWLNIQDIEKICQSSMNIHAIVHIATDYGRNRTPISEQYKCNVLLPTRLLELM
PALKTKFFISTDSFFGKYEKHYGYMRSYMASKRHFVELSKIYVEEHPDVCFINLRLEH
VYGERDKAGKIIPYVIKKMKNNEDIDCTIARQKRDFIYIDDVVSAYLKILKEGFNAGH
YDVEVGTGKSIELKEVFEIIKKETHSSSKINYGAVAMRDDEIMESHANTSFLTRLGWS
AEFSIEKGVKKMLSMKE"
CDS 1019. .2359
/note="wzx
Referred to as orf12.6 in reference [2]"
/citation=[2]
/codon_start=1
/transl_table=11
/product="hypothetical protein with 12 predicted
transmembrane segments which possibly functions in export
of O-antigen"
/protein_id="CAA43919.1"
/db_xref="GI:47006"
/db_xref="UniProt/TrEMBL:Q00339"
/translation="MNRIIRMLGVDKAIRYVIFGKIISVLTGLLLIMLISHHLSKDAQ
GYYYTFNSVVALQIIFELGLSTVIIQFASHEMSALKYDYSERDIIGESKNKQRYLSLF
RLAIKWYAVIALLIILIVGPIGYVFFTQKEGLGVPWQGAWLLLTIVTAFNIFLVSVLS
VAEGSGLITDVNKMRMYQSLLAGILAVSLLISGFGLYATSAIAISGTIIFSIFSYKYF
KKIFLQSLKHKNKYTEGGISWVNEIFPMQWRIALSWMSGYFIYFVMTPIAFKYFGAIY
AGQLGMSLTLCNMVMATGLAWISTKYPKWGVMVSNKQLAELSKSFKSAVMQSSFFVLT
GLTGVYISLWLLKLSGSNIGERFLGLQDFFFLSLAIIGNHIVACFATYIRAHKTEKMT
LASCIMALLTITTMLFVAYLEYSRFYMLMYAALTWLYFVPQTYIIFKRFKSSYE"
CDS 2352. .3314
/note="unnamed protein product; wbaR"
/codon_start=1
/transl_table=11
/protein_id="CAA43909.1"
/db_xref="GI:47007"
/db_xref="UniProt/TrEMBL:Q00337"
/translation="MSKKPLLTIAIPTYNRSSCLARLLDSIIQQENYCHDELEVIVCD
NASTDETARIAKSGLDKIRNSTYHLNEENLGMDGNFQKCFELSNGKYLWMIGDDDLIV
KNGISKVFSILKSRPALDMVYVNSAAKTELNYNADVRTSFYTNDVDFISDVKVMFTFI
SGMICKKTDAIVKAVGIFSPQTTGKYLMHLTWQLPLLKQGGEFAVIHNNIIEAEPDNS
GGYHLYKVFSNNLATIFDVFYPREHRVSKRVRASACLFLLNFIGDEDKTKNFATNNYL
RDCDSAFIDLIIYKYGLRFFYLYPKTVPLFRKIKYIIKTVLMRK"
CDS 3361. .3975
/note="unnamed protein product; wbaL
orf14.9 possibly encodes an acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="CAA43910.1"
/db_xref="GI:47008"
/db_xref="UniProt/TrEMBL:Q00480"
/translation="MFVYSLRLKLNLIISLLSKVRRKSKAKFLVLLSGYDFKMVGKNF
KLNVKPYSAKNNTSSKWGSMRVGDNCWIEAVYNYGDEKFEPYLYIGDRICLSDNVHIS
CVSCLILENDILIGSKVYIGDHSHGSYKVCSPKIEPPANKPLGDIAPIKIGNCCWIGD
NAVILAGSEICDGCVIAANSVVKDLKVDKPCLIGGVPAKVIKVF"
CDS 3977. .5020
/note="wbaQ
Referred to as orf15.6 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="rhamnosyl transferase"
/protein_id="CAA43911.1"
/db_xref="GI:47009"
/db_xref="GOA:Q00482"
/db_xref="UniProt/TrEMBL:Q00482"
/translation="MNVFISICIPSYNRAEFLEPLLDSIYNQDYCLKNNDFEVIVCED
KSPQRDEINSIIENYKAKNNKQNLYVNFNEDNLGYDKNLKKCISLTTGKYCMIMGNDD
LLADGALSKIVKVLKANPEIVLATRAYGWFKENPNELCDTVRHLTDDTLFQPGADAIK
FFFRRVGVISGFIVNAEKAKKLSSDLFDGRLYYQMYLAGMLMAEGQGYYFSDVMTLSR
DTEAPDFGNAGTEKGVFTPGGYKPEGRIHMVEGLLLIAKYIEDTTKIDGVYAGIRKDL
ANYFYPYIRDQLDLPLYTYIKMINKFRKMGFSNEKLFYVHAFLGYVLKRRGYDALIKY
IRSKKGGTPRLGI"
gene 5114. .6313
/gene="wzy"
CDS 5114. .6313
/gene="wzy"
/codon_start=1
/transl_table=11
/product="O antigen polymerase"
/protein_id="CAA43912.1"
/db_xref="GI:47010"
/db_xref="GOA:Q00474"
/db_xref="UniProt/Swiss-Prot:Q00474"
/translation="MLPFPPGAILRDVLNVFFVALVLVRFVIDRKKTYFPLVFTIFSW
SAVILWVIALTIFSPDKIQAIMGGRSYILFPAVFIALVILKVSYPQSLNIEKIVCYII
FLMFMVATISIIDVLMNGEFIKLLGYDEHYAGEQLNLINSYDGMVRATGGFSDALNFG
YMLTLGVLLCMECFSQGYKRLLMLIISFVLFIAICMSLTRGAILVAALIYALYIISNR
KMLFCGITLFVIIPVLAISTNIFDNYTEILIGRFTDSSQASRGSTQGRIDMAINSLN
FLSEHPSGIGLGTQGSGNMLSVKDNRLNTDNYFFWIALETGIIGLIINIIYLASQFYS
STLLNRIYGSHCSNMHYRLYFLFGSIYFISAALSSAPSSSTFSIYYWTVLALIPFLKL
TNRRCTR"
CDS 6313. .7323
/note="wbaW
Referred to as orf17.9 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="second mannosyl transferase"
/protein_id="CAA43913.1"
/db_xref="GI:47011"
/db_xref="GOA:Q00481"
/db_xref="UniProt/TrEMBL:Q00481"
/translation="MNNKKVLMDISWSNKGGIGRFTDEISKLLCDISKEELYRKCASP
LAPLGLAVNIFLRKKTDVVFLPGYIPPLFCSKKFIITIHDLNHLDLNDNSSLFKRLFY
NFIIKRGCRKAYKIFTVSNFSKERIVAWSGVNPNKIVTVYNGVSSLFNADVKPLNLGY
KYLLCVGNRKTHKNEKCVISAFAKADIDPSIKLVFTGNPCNDLEKLIIQHGLSERVKF
FGFVSEKDLPSLYKGSLGLVFPSLYEGFGLPVVEGMACGIPVLTSLTSSLPEVAGDAA
ILVDPLSEDAITKGISRLINDSELRKHLIQKGLLRAKRFNWQNVVSEIEMVLTEACDG
NK"
CDS 7310. .8467
/note="wbaZ
Referred to as orf18.9 in reference [2]"
/citation=[3]
/codon_start=1
/transl_table=11
/product="first mannosyl transferase"
/protein_id="CAA43914.1"
/db_xref="GI:47012"
/db_xref="GOA:Q00483"
/db_xref="UniProt/TrEMBL:Q00483"
/translation="MEINEIKISLVHEWLLSYAGSEQVSSAILHVFPEAKLYSVVDFL
TDEQRRHFLGKYATTTFIQNLPKAKKFYQKYLPLMPLAIEQLDLSDANIIISSAHSVA
KGVISGPDQLHISYVHSPIRYAWDLQHQYLNESNLNKGIKGWLAKWLLHKIRIWDSRT
ANGVDHFIANSQYIARRIKKVYRREASVIYPPVDVDNFEVKNEKQDYYFTASRMVPYK
RIDLIVEAFSKMPEKKLVVIGDGPEMKKIKSKATDNIKLLGYQSFPVLKEYMQSARAF
VFAAEEDFGIIPVEAQACGTPVIAFGKGGALETVRPLGVEEEPTGIFFKEQNIASLHEA
VSEFEKNASFFTSQACRKNAEKFSRSRFEQEFKNFVNEKWNLFKTEQIIKR"
gene 8482. .9903
/gene="manC"
CDS 8482. .9903
/gene="manC"
/codon_start=1
/transl_table=11
/product="mannose-1-phosphate guanyltransferase"

```
                     /protein_id="CAA43915.1"
                     /db_xref="GI:47013"
                     /db_xref="GOA:Q00473"
                     /db_xref="UniProt/Swiss-Prot:Q00473"
                     /translation="MSKLIPVIMAGGIGSRLWPLSREEHPKQFLSVDGELSMLQNTIK
                     RLTPLLAGEPLVICNDSHRFLVAEQLRAINKLANNIILEPVGRNTAPAIALAAFCSLQ
                     NVVDEDPLLLVLAADHVIRDEKVFLKAINHAEFFATQGKLVTFGIVPTQAETGYGYIC
                     RGEAIGEDAFSVAEFVEKPDFDTARHYVESEKYYWNSGMFLFRASSYLQELKDLSPDI
                     YQACENAVGSINPDLDFIRIDKEAFAMCPSDSIDYAVMEHTRHAVVVPMNAGWSDVGS
                     WSSLWDISKKDPQRNVLHGDIFAYNSKDNYIYSEKSFISTIGVNNLVIVQTADALLVS
                     DKDSVQDVKKVVDYLKANNRNEHKKHLEVFRPWGKFSVIHSGDNYLVKRITVKPGAKF
                     AAQMHLHRAEHWIVVSGTACITKGEEIFTISENESTFIPANTVHTLKNPATIPLELIE
                     IQSGTYLAEDDIIRLEKHSGYLE"
     gene            9911. .11347
                     /gene="manB"

  Query Match           100.0%;  Score 1341;  DB 1;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 4e-216;
  Matches 1341;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 1078

Qy         61 AAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 AAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCT 1138

Qy        121 AAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 AAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTT 1198

Qy        181 GAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 GAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAA 1258

Qy        241 TATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 TATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCG 1318

Qy        301 TTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 TTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTC 1378

Qy        361 GGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 GGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGG 1438

Qy        421 GCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 GCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCT 1498

Qy        481 GTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 GTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTG 1558

Qy        541 TTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 TTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCT 1618

Qy        601 GCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1619 GCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAA 1678

Qy        661 ATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 ATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTT 1738

Qy        721 AATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 AATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATT 1798

Qy        781 TATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 TATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTA 1858

Qy        841 GGGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 GGGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACT 1918

Qy        901 AAATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1919 AAATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCG 1978

Qy        961 TTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1979 TTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATAC 2038

Qy       1021 ATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2039 ATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTG 2098

Qy       1081 CAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2099 CAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCA 2158

Qy       1141 ACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2159 ACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTC 2218

Qy       1201 TTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACATGTTA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2219 TTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACATGTTA 2278

Qy       1261 ATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTTAAAAGA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2279 ATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTTAAAAGA 2338

Qy       1321 TTCAAGAGTTCTTATGAGTAA 1341
              |||||||||||||||||||||
Db       2339 TTCAAGAGTTCTTATGAGTAA 2359


RESULT 2
BD085556
LOCUS       BD085556               12441 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085556
VERSION     BD085556.1  GI:22631166
KEYWORDS    JP 2001522248-A/3.
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 12441)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 3 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Salmonella enterica
            PN   JP 2001522248-A/3
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .12441
            FT                   /organism='Salmonella enterica'.
FEATURES             Location/Qualifiers
     source          1. .12441
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:28901"
ORIGIN

  Query Match           100.0%;  Score 1341;  DB 6;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 4e-216;
  Matches 1341;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 1078

Qy         61 AAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 AAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCT 1138

Qy        121 AAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 AAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTT 1198

Qy        181 GAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 GAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAA 1258

Qy        241 TATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 TATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCG 1318

Qy        301 TTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 TTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTC 1378

Qy        361 GGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 GGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGG 1438

Qy        421 GCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 GCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCT 1498

Qy        481 GTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 GTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTG 1558

Qy        541 TTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 TTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCT 1618

Qy        601 GCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1619 GCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAA 1678

Qy        661 ATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 ATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTT 1738

Qy        721 AATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 AATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATT 1798

Qy        781 TATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 TATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTA 1858

Qy        841 GGGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 GGGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACT 1918

Qy        901 AAATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1919 AAATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCG 1978

Qy        961 TTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1979 TTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATAC 2038

Qy       1021 ATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2039 ATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTG 2098

Qy       1081 CAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2099 CAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCA 2158

Qy       1141 ACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2159 ACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTC 2218

Qy       1201 TTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACATGTTA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2219 TTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACATGTTA 2278

Qy       1261 ATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTTAAAAGA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2279 ATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTTAAAAGA 2338

Qy       1321 TTCAAGAGTTCTTATGAGTAA 1341
              |||||||||||||||||||||
Db       2339 TTCAAGAGTTCTTATGAGTAA 2359


RESULT 3
AX029263
LOCUS       AX029263               12446 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 3 from Patent WO9850531.
ACCESSION   AX029263
VERSION     AX029263.1  GI:10190114
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 3 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .12446
                     /organism="Salmonella enterica"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match           100.0%;  Score 1341;  DB 6;  Length 12446;
  Best Local Similarity  100.0%;  Pred. No. 4e-216;
  Matches 1341;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 1078

Qy         61 AAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 AAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCT 1138

Qy        121 AAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 AAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTT 1198

Qy        181 GAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 GAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAA 1258

Qy        241 TATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1259 TATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCG 1318

Qy        301 TTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 TTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTC 1378

Qy        361 GGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 GGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGG 1438

Qy        421 GCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 GCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCT 1498

Qy        481 GTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 GTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTG 1558

Qy        541 TTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 TTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCT 1618

Qy        601 GCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1619 GCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAA 1678

Qy        661 ATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 ATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTT 1738

Qy        721 AATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 AATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATT 1798

Qy        781 TATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 TATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTA 1858

Qy        841 GGGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 GGGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACT 1918

Qy        901 AAATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1919 AAATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCG 1978

Qy        961 TTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1979 TTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATAC 2038

Qy       1021 ATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2039 ATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTG 2098

Qy       1081 CAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2099 CAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCA 2158

Qy       1141 ACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2159 ACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTC 2218

Qy       1201 TTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACATGTTA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2219 TTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACATGTTA 2278

Qy       1261 ATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTTAAAAGA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2279 ATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTTAAAAGA 2338

Qy       1321 TTCAAGAGTTCTTATGAGTAA 1341
              |||||||||||||||||||||
Db       2339 TTCAAGAGTTCTTATGAGTAA 2359
```

```
RESULT 4
CP000076_58
WPCOMMENT
  Sequence split into 71 fragments  LOCUS CP000076 Accession CP000076
     Fragment Name      Begin      End
     CP000076_00            1    110000
     CP000076_01       100001    210000
     CP000076_02       200001    310000
     CP000076_03       300001    410000
     CP000076_04       400001    510000
     CP000076_05       500001    610000
     CP000076_06       600001    710000
     CP000076_07       700001    810000
     CP000076_08       800001    910000
     CP000076_09       900001   1010000
     CP000076_10      1000001   1110000
     CP000076_11      1100001   1210000
     CP000076_12      1200001   1310000
     CP000076_13      1300001   1410000
     CP000076_14      1400001   1510000
     CP000076_15      1500001   1610000
     CP000076_16      1600001   1710000
     CP000076_17      1700001   1810000
     CP000076_18      1800001   1910000
     CP000076_19      1900001   2010000
     CP000076_20      2000001   2110000
     CP000076_21      2100001   2210000
     CP000076_22      2200001   2310000
     CP000076_23      2300001   2410000
     CP000076_24      2400001   2510000
     CP000076_25      2500001   2610000
     CP000076_26      2600001   2710000
     CP000076_27      2700001   2810000
     CP000076_28      2800001   2910000
     CP000076_29      2900001   3010000
     CP000076_30      3000001   3110000
     CP000076_31      3100001   3210000
     CP000076_32      3200001   3310000
     CP000076_33      3300001   3410000
     CP000076_34      3400001   3510000
     CP000076_35      3500001   3610000
     CP000076_36      3600001   3710000
     CP000076_37      3700001   3810000
     CP000076_38      3800001   3910000
     CP000076_39      3900001   4010000
     CP000076_40      4000001   4110000
     CP000076_41      4100001   4210000
     CP000076_42      4200001   4310000
     CP000076_43      4300001   4410000
     CP000076_44      4400001   4510000
     CP000076_45      4500001   4610000
     CP000076_46      4600001   4710000
     CP000076_47      4700001   4810000
     CP000076_48      4800001   4910000
     CP000076_49      4900001   5010000
     CP000076_50      5000001   5110000
     CP000076_51      5100001   5210000
     CP000076_52      5200001   5310000
     CP000076_53      5300001   5410000
     CP000076_54      5400001   5510000
     CP000076_55      5500001   5610000
     CP000076_56      5600001   5710000
     CP000076_57      5700001   5810000
     CP000076_58      5800001   5910000
     CP000076_59      5900001   6010000
     CP000076_60      6000001   6110000
     CP000076_61      6100001   6210000
     CP000076_62      6200001   6310000
     CP000076_63      6300001   6410000
     CP000076_64      6400001   6510000
     CP000076_65      6500001   6610000
```

```
    CP000076_66       6600001   6710000
    CP000076_67       6700001   6810000
    CP000076_68       6800001   6910000
    CP000076_69       6900001   7010000
    CP000076_70       7000001   7074893
Continuation (59 of 71) of CP000076 from base 5800001 (CP000076 Pseudomonas fluorescens

  Query Match             6.8%;  Score 91;  DB 1;  Length 110000;
  Best Local Similarity   45.1%;  Pred. No. 1.8e-06;
  Matches  512;  Conservative    0;  Mismatches  605;  Indels   18;  Gaps    4;

Qy         32 ATAAAGCAATTCGTTATGTTATTTTTGGTAAGATAATATCTGTATTAACGGGTTTACTGT 91
              ||||||   |||||  |       | |   | |||||       | ||  | ||     | |
Db      61047 ATAAAGACTTTCGTCACACAGCTATAAATCAGATATGGCGACTCTTTTCCGGACCTTTAT 61106

Qy         92 TAATAATGTTAATATCACACCATTTATCTAAAGACGCACAGGGCTATTATTATACATTTA 151
              |  |  |  |  |  |     |||||  |   ||| |||||  || ||||  |||||  ||
Db      61107 TGCTTGTACTTCTGCCGTTATATTTAACACCAGAGGCACAAGGTTATTGGTATACCTTCG 61166

Qy        152 ATTCAGTAGTGGCACTACAGATAATATTTGAATTGGGGCTATCAACGGTAATCATTCAAT 211
              |   | |  ||  |   |||  |   |||  ||||  | ||  || || |  ||  ||
Db      61167 TTAGCCTGGCCGCCTTGGCAATATTTGCTGACATGGGTTTTTCCGCGATATTACTTTTAT 61226

Qy        212 TCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTATTCTGAACGAGATATTATAGGTG 271
              || ||   |||||| |  |     | || |              ||  |   || |  || |
Db      61227 TCTCTGCACATGAATTTGCGCATCTTAACTTCAGAAGCGACAAAACA---ATAACCGGAG 61283

Qy        272 AAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCGGTTGGCAATAAAATGGTATGCAG 331
              |     |  |   |  |||          |  | | || ||  ||   || |||| | |
Db      61284 ACGAGTACCACCTGATACGCCTGGCGACACTTCTGCGTTTTTCAGCGAAGTGGTCTTCGC 61343

Qy        332 TAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGTATGTTTTTTTTACGCAAA 391
              | || |||||| |  || |    ||||||      || || |||    |||| |     |
Db      61344 TCATGGCTTTGATTGTATTCCCTATAGTCTTGATTATTGGCTATACGATTTTAAACGCGA 61403

Qy        392 AAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGGTTATTATTAACAATAGTTACAGCTT 451
              |||        |  |  |||| |    ||||||  |  | |       |   |||  |
Db      61404 GAGA---TACTACTACAGATTGGAAGCTTGCATGGCTCATCTATGGGGCTGCATCAGTCT 61460

Qy        452 TTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGGTTAATTACTGATG 511
              |     |  |          || || ||  ||  ||   |||||  ||  |    |    || |
Db      61461 TCGTGTTCATCAACAGCATGGTGCTATCGTTCATCGAAGGTTGTAACTCGGTAGGAGACG 61520

Qy        512 TGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATATTGGCAGTAAGCTTACTTA 571
              |  | |||||  |  |  ||     || |  |   |||     |  ||| | |||||
Db      61521 TTCAGAAAATTCGTTTTCATATAAGTGTGGTGACATTTATCACTACTCTAATCCTACTTG 61580

Qy        572 TTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTTCAGGGACTATCATATTCT 631
              |||  ||     | || || ||    ||          |  | |||  |  |
Db      61581 TTACGGGGGCAAGTCTCTACGCCCTCTCACTCTCGCTGCTAGCGGGGGCGTTGGCTGGGA 61640

Qy        632 CCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAAGCATAAAAATA 691
              |   | |             | |  || | |||  |     |||   |          | | |
Db      61641 CTGCACTCATCACGTTCCACTACAACAGAATGCTGTATCAACTCT---------ATCAAA 61691

Qy        692 AATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCTATGCAATGGCGAATTG 751
              ||   || | ||     |  | | ||| |     |||||  | ||  ||  ||||||  |||
Db      61692 AAGCAACAGGAGAATCTCATCCCTGGATGCGGGAAATCATGCCATTGATATGGCGCTTTG 61751

Qy        752 CTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTATGACCCCCATTGCATTCAAAT 811
              |  |||| |||||     || || ||||| | ||    |  | || ||  | |||||  | |
Db      61752 CAATAAGCTGGATCAGCGGCTACTTTATATTTTCCATATTTACTCCTCTAGCATTTCATT 61811

Qy        812 ATTTCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTTTAACATTATGCAATATGGTAA 871
              | | |||  | ||  | |||||  |  | ||  | ||| |  |  |||         ||
Db      61812 ACTACGGCACGATTGACGCAGGAAAAGTTGGTCTCTCTATCGCGGTCTGCCTAGCCATAT 61871

Qy        872 TGGCTACGGGCCTGGCTTGGATATCCACTAAATATCCAAAATGGGGAGTAATGGTTTCCA 931
              |    |  |       |||||  | |  | |   ||   |  |     |    || |
Db      61872 TCAGCATCGCAAACATCTGGATCACGATAATAACCCCTCGAATGAACATCTATGTCGCGC 61931

Qy        932 ACAAACAGCTTGCGGAACTGAGTAAATCGTTCAAAAGTGCAGTAATGCAATCATCCTTTT 991
              | ||  ||   |  |||| |        || || |     |  | |  |   | |
Db      61932 AAAAGGAGTACGTCAAACTTAACTCCGACTTTAACAAGCACGCTCTTCTTGCGATCATGA 61991

Qy        992 TTGTCTTGACAGGATTAACTGGTGTATACATTTCATTATGGTTATTGAAATTATCTGGTT 1051
                  |   ||  |||| | | |||||||  ||||  | || ||  ||   |
Db      61992 CCTATATCTTGGGGATAACAGCTCTATACATAGCATTGAGTTTTTTTCAAGACTACCTGC 62051

Qy       1052 CAAACATTGGCGAGCGGTTTTTGGGATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTA 1111
              |   |||   || ||  |  ||      |      ||| |  || |  | || ||   |
Db      62052 C---GATTCTAGACCGCCTACTGCCGACGGCTTCCTTTCTTATTATTGCCTTGGCCTATC 62108

Qy       1112 TTGGTAATCACATTGTAGCTTGCTTTGCAACCTATATAAGAGCGCATAAAACTGA 1166
              ||| | | |  |    |  | | |||   ||||| ||||| || ||||  ||
Db      62109 TTGCTCAACTTACCATTAACTCCATGGCAGTTTATATGAGAGCTCAAAAAAAAGA 62163


RESULT 5
PFA929358/c
LOCUS       PFA929358             254050 bp    DNA     linear   INV 16-APR-2005
DEFINITION  Plasmodium falciparum strain 3D7, chromosome 9; segment 4/5.
ACCESSION   AL929358 AL844508
VERSION     AL929358.1  GI:23505147
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D.,
            Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K.,
            Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C.,
            Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C.,
            Cronin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J.,
            Feltwell,T., Goble,A., Goodhead,I., Gwilliam,R., Hamlin,N.,
            Hance,Z., Harper,D., Hauser,H., Hornsby,T., Holroyd,S.,
            Horrocks,P., Humphray,S., Jagels,K., James,K.D., Johnson,D.,
            Kerhornou,A., Knights,A., Konfortov,B., Kyes,S., Larke,N.,
            Lawson,D., Lennard,N., Line,A., Maddison,M., McLean,J., Mooney,P.,
            Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A.,
            Rabbinowitsch,E., Rajandream,M.A., Rutter,S., Rutherford,K.M.,
            Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R.,
            Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L.,
            Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and
            Barrell,B.G.
  TITLE     Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
  JOURNAL   Nature 419 (6906), 527-531 (2002)
   PUBMED   12368867
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Sanders,M., Hauser,H., Baker,S., Unwin,L., Mungall,K., Berriman,M.,
            Pain,A., Hall,N., Bowman,S., Churcher,C., Quail,M. and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2002) P.falciparum Genome Sequencing Consortium,
            The Welcome Trust Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .254050
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="9"
     gene            763. .2532
                     /gene="PFI1215w"
     CDS             763. .2532
                     /gene="PFI1215w"
                     /note="Moderately similar to Mus musculus splicing factor
                     3a, subunit 3, 60kd TR:AAH09141 (EMBL:BC009141) (501 aa)
                     fasta scores: E(): 1.7e-24, 30.756% id in 595 aa. No
                     significant pfam hit; SMART hit to SM00355, zinc
                     finger;ProfileScan hit to PS50171, Zn-finger (C2H2) in
                     matrin and other RNA-binding proteins"
                     /codon_start=1
                     /product="splicesome-associated protein, putative"
```

/protein_id="CAD51929.1"
/db_xref="GI:23505148"
/db_xref="GOA:Q8I2R0"
/db_xref="InterPro:IPR000690"
/db_xref="InterPro:IPR007087"
/db_xref="UniProt/TrEMBL:Q8I2R0"
/translation="MSQFLIEQIRYLHEEIEVIEKAIAELIDEKVKNKKKNILYDYAI
NYLVEKIQEKSQLLLEYYNDEDKLKKEEMLFLSGKTGEDKNDIWKNYYERVKYIRDYH
KKTNIKKIEIKSSKLYKLEALKCSKYKQSFSPLEKKGKYVDLNKFYNDFINMKKIKEF
RISMFKKKEMASSHNRKRKKTGDNNNNNNNNNNVTTANINEFKEMDYVTYLHNFTRFI
YIPRYCKYRNEEYKSYVENLLNYMISFFEKINVLVDCKTTYKEYEDDFYNKFNNKEIK
GWEKYTYDLDFYCNINNKLYASEGTYQSYLKCKNYDNDLKKYLNKKYTVEQLEEIKKN
IEKEDMNLARCEYLIQLYANILNKIIQRTIQNIQRKQAYTIDEIQKKKYKEEEQEDEK
KRKKEYHLLSLMDHNNNELNNQNILLQDLENNNLSISMSDDDSSDMSNISNNEDQQKD
LDEENEDDNDDNKPIYNPLNLPLGFDNKPIPYWLYKLHGLSKEYKCEICGNYSYFGRA
TFEKHFYEWRHSFGMKCLKIPNTLHFKEITKIEDALNLYEKLKKETQMNIFKPDQEVE
CEDSKGNVMNIKAYDDLKRQGLL"
gene join(4370. .4451,4555. .4604,4733. .4964,5141. .5186,
5417. .5444,5735. .5773,5790. .5825)
/gene="PFI1220w"
CDS join(4370. .4451,4555. .4604,4733. .4964,5141. .5186,
5417. .5444,5735. .5773,5790. .5825)
/gene="PFI1220w"
/note="1 probable transmembrane helix predicted for
PFI1220w by TMHMM2.0 at aa 74-96"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51930.1"
/db_xref="GI:23505149"
/db_xref="UniProt/TrEMBL:Q8I2Q9"
/translation="METFNHIDIKKYTREDLLYLRNTNNVLFKMFQKQVDEIACLSSK
EQKLCGTLTSINNIINENYTIYVNTKKEHIIYIYVYVYVFICLCICSFIYFYMYLILS
CQVLNTHRWLNRIYKSNNKLYMCVRLKIKKMKGSTQRHFCFLFMPNHFVIFSNYFKNY
MREFLINMCI"
gene join(6108. .6182,6392. .6547,6648. .6699,6851. .6903,
7009. .7062,7151. .7204,7343. .7424,7603. .7710,7795. .7853)
/gene="PFI1225w"
CDS join(6108. .6182,6392. .6547,6648. .6699,6851. .6903,
7009. .7062,7151. .7204,7343. .7424,7603. .7710,7795. .7853)
/gene="PFI1225w"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51931.1"
/db_xref="GI:23505150"
/db_xref="UniProt/TrEMBL:Q8I2Q8"
/translation="MNEKEKNGIFSVHKEKDNLSTHQKFDLLIDEYYCDLKEEKRERS
TEKNTKEGDFVFKSVDRNKADYAVRNRRKMLHHRLDENDEPPIFTDLIENEKKIGTVR
ITNPLPKIERRDELKDNYMKWLKTERLMRPNERRLEYEQMQEKYLESIKLTHDIKKDV
KLAKAIKSHYPRGIVGVDSIYNENTVLYKDQHEEIKKKYEIKEKRLKEKGEVLEKHNQ
KNGNFLTFQENN"
gene complement(8468. .9508)
/gene="PFI1230c"
CDS complement(8468. .9508)
/gene="PFI1230c"
/note="asparagine-rich
Signal anchor predicted for PFI1230c by SignalP 2.0 HMM
(Signal peptide probabilty 0.001, signal anchor
probability 0.995) with cleavage site probability 0.001
between residues 27 and 28
4 probable transmembrane helices predicted for PFI1230c by
TMHMM2.0 at aa 12-34, 38-56, 63-82 and 97-116"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51932.1"
/db_xref="GI:23505151"
/db_xref="UniProt/TrEMBL:Q8I2Q7"
/translation="MEERKWSYIMVSVIFYLCYFFYFLTYSYEIIFLFQTKIDYIVFC
LLFITSLFSIYFCSKGLKYICTFLITIISLIEIYEYTNNIPKDLYDEEFLKLFLLVRL
FSAVNLVVFLYLHIKNLKNIKDEKLNNDKDSTIQSVAEPTNTTTENNHNQNNNHNNNN
IHNNNHNNNHNNHNNIHNNHNDPCNNHNDPCNNYYKNQDNFLKSNTVTYSHNNLFKEN
IPHVHQHTGIITSKTLPQENIYVLKHVSIPIFEKEIINKDTNISHMYEMQTNDSFNHY
QVNNLSSSNNINDNKITEPITTHNLNKNNNAYTNHIIQNNHVPNYTYEHPKGLQYYPG
NFHTHNMGNNIQ"

gene join(11425. .12434,12559. .12655,12774. .12908)
/gene="PFI1235w"
CDS join(11425. .12434,12559. .12655,12774. .12908)
/gene="PFI1235w"
/note="No pfam hit; ScanProsite hit to PS00430; Interpro
hit to Ton-B dependant receptor protein"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51933.1"
/db_xref="GI:23505152"
/db_xref="InterPro:IPR007823"
/db_xref="InterPro:IPR010916"
/db_xref="UniProt/TrEMBL:Q8I2Q6"
/translation="MTDINNNKYSKLKVKKNKNKKNNNKNNDKNNTLLVSPNDKLKKE
KRKKKHKNLNQKVIHNNNNNNNIHKDEHINEQSNNHIYVNNRTKKKKKNIELIKDKWI
NKKKKKKTKKKTKKKTNNTTIYDNDNNINVDEEEEKKKKNVFPYISVLRENVVNNKIH
IHDTNEAYKYDYADLENNKQKYKKKKKIENSPEDIVNSSLFRYINEYMYTNNSEVVQN
KLNQTKNVFNIYHQGYKNQKNKWPHNPVSIIIKHLKKYFNKNNKIADLGCGEAEIART
LDGWYIKSFDLIQYNHYVTPCNITQLPLNNNSYDCFILSLSLMNTDWPKIIFESVRCL
KKGATLIIAEVVSRFTNYKAFIKFMNNVGFKLSNKINLDDFFYVFFFEKNQEVDISSS
INEKRIKKVSSLLTPCIYKRR"
gene complement(13154. .14893)
/gene="PFI1240c"
CDS complement(13154. .14893)
/gene="PFI1240c"
/EC_number="6.1.1.15"
/note="Similarities to prolyl-t-RNA synthases of bacterial
origin; N-terminal signal peptide. Similar to Lactococcus
lactis prolyl-trna synthetase pros or ll2127 SWALL:Q9CDT4
(EMBL:AE006441) (616 aa) fasta scores: E(): 1.6e-09,
35.21% id in 213 aa, and to Escherichia coli prolyl-trna
synthetase pros or drpa or b0194 SWALL:SYP_ECOLI
(SWALL:P16659) (572 aa) fasta scores: E(): 7.3e-08, 31.91%
id in 235 aa. Pfam match to PF00587, tRNA synthetase class
II (G, H, P, S and T)
Signal peptide predicted for PFI1240c by SignalP 2.0 HMM
(Signal peptide probabilty 0.720, signal anchor
probability 0.018) with cleavage site probability 0.591
between residues 18 and 19"
/codon_start=1
/product="prolyl-t-RNA synthase, putative"
/protein_id="CAD51934.1"
/db_xref="GI:23505153"
/db_xref="GOA:Q8I2Q5"
/db_xref="InterPro:IPR002314"
/db_xref="InterPro:IPR004154"
/db_xref="InterPro:IPR006195"
/db_xref="UniProt/TrEMBL:Q8I2Q5"
/translation="MLIKIIILIGLFHHVILSLKLVNLKKKRSNVFFVNNKILKRNVP
KKKYRILSNKNQSYIGSNYIFNRIFNDNIKSNGDKTSELLQRANYIRDVNGIYNILPL
GFRVINKIIDFLNKHLEKLNSHAMSLSILQAKKLWNISDRSKLYSDEFLYVYKQKQSK
NDKSSEESLKKKIEDDGYILSPTCEESSLSLINQIYNENITIKCLPLLIHQYNYKFRN
EKRFEKSLFKSKEFLMKDGYSFHSNEKCLNETYEKYKECYKNIFEELKLSFNIIKKRK
KDKMNALESHEFQVLSRDGKYKEAAHIFKLGDYYSNKLDIKYLDKKNEKKNILMGSYG
IGIYRLLYFLIENFYDEEGIKLPQQVAPFSVYLIQTNQKSKYSATKISKILNMYKDSM
EKKGNISQQKNEQDDQGDDECGNSTFVSNIHNDDKVNKDDTLNNNNDTLNSNNTLNHD
NGVNINNTLTYSPLTSNDIEYILTLWLFNTFKNKNVDIYYDDTDLHLSRKLKHCDLIG
APNRIIINLSNLDKKVKVPTNFYNYIDNSDNILLNKLYLTFQNITIEYKHRFSQEKKI
MTIRELFKYFKFI"
gene complement(join(15848. .16581,16944. .17610))
/gene="PFI1245c"
CDS complement(join(15848. .16581,16944. .17610))
/gene="PFI1245c"
/note="Plasmodium falciparum protein phosphatase-beta
TR:O15920 (EMBL:U89025) (466 aa) fasta scores: E():
1.9e-180, 99.142% id in 466 aa. Pfam match to PF00149,
Ser/Thr protein phosphatase; SMART hit to SM00156, Protein
phosphatase 2A homologues, catalytic domain."
/codon_start=1
/product="Protein phosphatase-beta"
/protein_id="CAD51935.1"

Query Match 4.8%; Score 64.4; DB 2; Length 254050;

```
  Best Local Similarity   43.3%;  Pred. No. 0.037;
  Matches  350;  Conservative    0;  Mismatches  456;  Indels    2;  Gaps    1;

Qy          9 TATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGTAAGATAAT 68
              ||| |||| |||       |  | ||||  |||   |||| ||  |||   |||    ||
Db     237191 TATAATTAAAATTAATAATAAAATTAAATAAATATATTATTTTTATTTATTTAATTATAT 237132

Qy         69 ATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCTAAAGACGC 128
              || | |||       |||| | | | |||  | |||||      ||| ||  ||| |
Db     237131 ATATATATATTATTTTTTATTATAATTAAAATAAATATTTATTAATTAATAAAAAAATAT 237072

Qy        129 ACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTTGAATTGGG 188
              |      |||    |  | ||||| ||          | |  | |||||||  | ||||
Db     237071 ATTAATGTATATATTAATATTTATTTATTATTCATTAATTAAAATAATATATTAATTTTA 237012

Qy        189 GCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTA 248
                |||  |   ||    |   |           | | || |      |  | |  | | ||
Db     237011 CACATCCATAATATATTTAATAAATTAATATATTTATATTTATATATATATAATTAAATA 236952

Qy        249 TTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCG 308
              | | |   | ||  |  |  |  ||  || | |   ||  |    ||   ||  ||
Db     236951 ATTTAATTTATATTATTAAATTATAATGAATATAATTAAATATATAATAAATTTTATTAA 236892

Qy        309 GTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAA--TATTAATAGTCGGTCCC 366
              |    |   ||    ||    || | | ||   ||||||  || | |||
Db     236891 ATATTAGAATTAAAATAATTATTTATTTGTTTATTTAATAAATTAATTATATATATAATA 236832

Qy        367 ATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGG 426
              ||    ||| | ||    |   |  ||  |    ||  | |   ||   ||    | |
Db     236831 ATAATATATTTATTAAATTTTTATTAACTATTAATATATATTATTTAATAATATTCCTTA 236772

Qy        427 TTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCT 486
              | || |||| |  |  || |   ||| || | ||| ||  || | |  ||| | | |
Db     236771 ATTTTTTTAAAATAATATAAAATGTTTTATTTATTTATTTATTTTTTTATTTATTTATTT 236712

Qy        487 GAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCT 546
              |   | |    ||    |   |  ||| | |    | | | | |   |  | |
Db     236711 TTATATATTTATATATAATATTTATTTTATATATTTTATGTATTTAAATTATTAAAATGT 236652

Qy        547 GGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATA 606
              |   |      | | | | | ||    | |  |      |    | ||
Db     236651 TATTATTATCTTATTATTTAATTAATATTATATTTATTAATTATTTAATAATTATAATTT 236592

Qy        607 GCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTC 666
              |  || |      |||   | |       ||||    |||  || ||||| | |  ||
Db     236591 ATTTATTAATTTTTTATTTCAATTAATTAATTTATTAATATTTAATTTAATTATAAATTA 236532

Qy        667 CTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAA 726
              |  |    ||||    |||  ||  |||||          | | | |     || |
Db     236531 TTTAATATATTAATTTTTAATTATTCATATATAATTAATTATTTATTTATTTAATTT 236472

Qy        727 ATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTT 786
              ||  ||  |||  |||     |||   ||  ||   ||| |    ||||||  |||||
Db     236471 ATTATTTTTATTTAATATTATATTTAATTACTTTATTTGTAATTTAATTTTAGCTATTTA 236412

Qy        787 GTTATGACCCCCATTGCATTCAAATATT 814
              |||| |      ||   ||| |  ||||
Db     236411 TTTATTATATATATATAATTAATTTATT 236384


RESULT 6
AC116956/c
LOCUS       AC116956               266544 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 1418423-1684967 strain
            AX4, complete sequence.
ACCESSION   AC116956 AC115678 AC116980
VERSION     AC116956.2  GI:28828496
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 266544)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 266544)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 266544)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 266544)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 4, 2003 this sequence version replaced
            gi:20042935, gi:19682974, gi:19919988.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .266544
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="1418423-1684967"
     CDS             join(<842. .1240,1322. .1916)
                     /note="GeneID exon scores (in order of location ranges):
                     29.76, 65.85 - GSCJ_ID dd_02843"
                     /codon_start=2
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Hypothetical 98.7 kDa protein (Fragment)"
                     /protein_id="AAO51167.1"
                     /db_xref="GI:28828559"
                     /translation="VAISNFNINNNNKNNNDNKEIRFPTDIIIDDSTEINFNNHEDNN
                     NNNNNNNNNNNAYNEDSGNSVQLGDIECVVCLDFLDSYLPTLVKIISEYGVIESCSKI
                     CGLLNKTDEIDICTGLCDLVGVDTFWKIFVTNDINPFYACELITACSTPKNPAADFIS
                     LGVSPAIGQSGSKFLIDLTFQVVNTTGAGQFAFIVFYTQTQSKFINYQLFEGFSPGNY
                     SVKYDFETTNNSTFINGQYPVTVYMCNGECGTMYSTLLNQTTSYFTISNPTPTPTPTP
                     SNSTTPTPTPTNSTPTPTSTSTPTSTPTSTPTPTPTSSSSTKTHSSHYKNKINK"
     CDS             complement(join(2119. .2662,2752. .2943,3024. .3055))
                     /note="GeneID exon scores (in order of location ranges):
                     32.18, 11.63, -1.19 - GSCJ_ID dd_02844"
                     /codon_start=1
                     /product="similar to Mus musculus (Mouse). 16 days neonate
                     thymus cDNA, RIKEN full-length enriched library,
                     clone:A130067E11 product:cyclin C, full insert sequence"
                     /protein_id="AAO51105.1"
                     /db_xref="GI:28828497"
                     /translation="MAANFWESSHCTEWLLDRNKIEESNPKDKQYLTPIELKKLRTHY
                     CFVIQNLGNALKLRQRAISTAIVYFKRFYLKNSFVDCEPRLVAVTCLYLSSKVEECIT
                     QAKKCAAKMKEIDHSFNYLMNDILECEFFVLEELDFCLIIYHPYKSLPFYLQSSGLDP
                     ASIEIIWGIVNDSYRTDVCLLYPPFVVGLGCILLGSYLLKKDIKQWLSELNVEMKDIW
                     EVSKDLIDYYEFEKQQSLQNQSELDLLYNKLPIRNKK"
     CDS             complement(join(3639. .4158,4352. .4422))
                     /note="GeneID exon scores (in order of location ranges):
                     63.23, 9.00 - GSCJ_ID dd_02845"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
```

```
                     mold). Vacuolar ATP synthase proteolipid subunit (EC
                     3.6.1.34)"
                     /protein_id="AAO51106.1"
                     /db_xref="GI:28828498"
                     /translation="MFQLLSFLLSGEATAVERIITDACPVYAPFFGAMGVTAALVFTV
                     MGAAYGTAKASVGISNMGVMKPDLVIKAFIPVIFAGVIAIYGLIICVILVGGIKPNAN
                     YTLMKSFTDLGAGLTVGLCGLAAGMAIGIVGDSGVRAFGQQPKLYVIMMLILIFSEAL
                     GLYGLIIGILLSSVSDTYCPGQALVPLNSGNVIGKN"
CDS                  join(5902. .6004,6157. .6199,6351. .6486,6853. .6962,
                     7066. .7249)
                     /note="GeneID exon scores (in order of location ranges):
                     6.21, -0.71, 4.93, 17.98, 20.83 - GSCJ_ID dd_02846"
                     /codon_start=1
                     /product="similar to Mus musculus (Mouse). Similar to
                     ATPase, H+ transporting, lysosomal (Vacuolar proton pump)
                     21kD"
                     /protein_id="AAO51107.1"
                     /db_xref="GI:28828499"
                     /translation="MPDINPLTPIPDNNYYDQCTYFLVTISPSTWAALGIGLSLALSV
                     VGSAWGIWVTASSLMGAAVKEPRIRSKNIISIIFCEAVAIYGIILAIILNGKIDKFLN
                     IWDPASDYMAGYMMFGAGITVGLCNVFSGVCVGIAGSGCALGDAQNPSLFVKMLIIEI
                     FAGALGLYAVIVGILMTTNALGVTQQPALKP"
CDS                  complement(7573. .8283)
                     /note="GeneID exon scores (in order of location ranges):
                     66.20 - GSCJ_ID dd_02847"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO51108.1"
                     /db_xref="GI:28828500"
                     /translation="MSTTIYFSYPTYSNEREFTTYDDDSEEEYYDQSYFQDYSFQNKP
                     QPSSSNINKSKSVKVTGYGNGSKNAGQAIPRTDKDYKAAIEYKKEKKLAIKQIKENKL
                     STSPTTITTTTTTNTATSYENVNIESPNDENNKTKKLTKLEVKSIIEKETDLKKKERL
                     PSSIEVYNRILHDEQLKDKIQDFIITYEDRFDGLCDSELLKLDEAIPFHRIRQIRYLN
                     QIVWSRSERFYDFSNLLD"
CDS                  complement(join(9480. .10271,10379. .11125,11253. .11858,
                     11963. .12121))
                     /note="GeneID exon scores (in order of location ranges):
                     50.34, 56.92, 38.65, -2.17 - GSCJ_ID dd_03351"
                     /codon_start=1
                     /product="similar to Plasmodium falciparum (isolate 3D7).
                     Hypothetical protein"
                     /protein_id="AAO51109.1"
                     /db_xref="GI:28828501"
                     /translation="MEDTALMGYFHFIRYLHAATTPKLRFAFEIFSLIVISLFVSTLF
                     VMHNQFVGKSDCLKLPNNTRFDILSIEIKGETSIFSRQHIYKKSTTTTETTPTTTTTT
                     TTTVNNDIDQQSYNKYLQVLKGAVDKHEPCIYDENYDNDNDNDIDKDSVTTATTTTAA
                     AAAAAAAATTTTTTTSLITESTYFKTAIEFIIDLIRQMYHSLYNFISIFFNSQFDEYD
                     QQYKKDRLEREHLKRLKEELKKHIKDIDRNHVNFEKLQHHQQQQLQQPQQQQQQQQQQ
                     QQQQQQQQQQQQQQQQEREEKKEKDKQHNNNHHNNGIGSGVLNDLNDIVQYFIDSHNP
                     RYEFSYQNAFLMLTTEQRNKFNITTLHIDINANDQCLGNGFHKAIIGMLGYDILIINN
                     LVYGFPLIPNNNINNNNNNNNGRGGYLRIAGNSQLINLYQYSSPKLSLNNESILYIGY
                     RFFETNFIFFLTLIVFGSSLKLCESSFAHYVDRDPRGVLKCMLHNLLIFALFIFIQIG
                     YISIFSIILSTEYLVFFISSLITSYFSTWGLRSKESIRYYPIIFLSILYGLILYVVFF
                     PSGFHPIAFYACYSLIEYLLIICLFNFEIPAVLNNRILPNNQNNNNQNNNNQNNNNNN
                     QNNNQNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNQ
                     NNNQNNNNNQNNNNNNTNINNHNNNNINNQNNINNTQNENQNNNNQNDGNNQNIYNVP
                     INQQTIVNNSSQRRLAASAILNNSNEM"
CDS                  complement(join(13033. .14475,14560. .15048))
                     /note="GeneID exon scores (in order of location ranges):
                     138.82, 42.31 - GSCJ_ID dd_00640"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO51110.1"
                     /db_xref="GI:28828502"
                     /translation="MNIINYFKNGNNNNGHSSGVGSSGHSYYSNNKGTSFVPSPVLYM
                     TPMAPIDYGCPDDMLDDGRSDTAPSSQNTSPMYQHHNFHNNHNNTHNHNNNNNNNNNN
                     HEQPNCALEFLEHHLSFETKYGVNVSQGNRKYQEDRHKVKMGLENNQYLSLFGVFDGH
                     GGDRASNFVKKKIVNCVNKYVKENKAGYSSKNLNSSSSPTGSTSSASSSGSSSSSASS
                     SGGGGGGGGPLNGSSSSIGMPHASTISPTPSSGNSSTSSGASSSGVSSANSPIGVGFD
                     SPGWMKFNHLYNQNNFQQEIQNRSEFLQAALYNTFQFLDNRYCKKYRQKGDGGTTCLV
                     ALLSNPPNAQPLLVVANAGDSRGVLCRNGKAYALSYDHKPGNPKEKQRITSSGGKIEW
                     DFNERIWRVSGILSVSRGIGDIPLKKWVICDPEFVVFPLKGPIRKSNSSRYTSPSSSS
                     SSSSSSSSSSSSSPSSPKAFRSSPQINRFFFSPSFKNKTISNSTGSLPPPSTNNSPQ
                     QPPPILHQNNISSPVSSISEAQPFNSIDNQNNNNNNNNNSKLISTSPPNHKMYYCSNH
                     QYNYGEVDQFFVLATDGIWDVFENQELVEFINAIIEESYHSKRLDWDPNEISKRVVQE
                     AYRKGSGDNATVLIIKLYW"
CDS                  complement(join(17600. .18734,18858. .19218,19315. .20422))
                     /note="GeneID exon scores (in order of location ranges):
                     69.94, 30.45, 124.05 - GSCJ_ID dd_00639"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). cGMP-stimulated cGMP phosphodiesterase
                     (cGMP-stimulated cGMP-specific phosphodiesterase)"
                     /protein_id="AAO51111.1"
                     /db_xref="GI:28828503"
                     /translation="MMFSKRKSIRFNPEGDYTELPRGGYVVPSKLGGIQFGVPPETIK
                     DSMALKIDVPTIYVFPEELWDRKTGINAAEAEFPAYFNYFILKRKVSFVCTKEQEQRI
                     RIVFQETLLGPPEFNTGIIINSSSSTTDTSKTSPIKKQTSSSSPPLSPQQQQPPPPLV
                     KQPSQQQLEELATMDTCHYHHHHHHQEVNDNDNNNNTTTNNNNIEILEQQQQQQQQQQ
                     QQQDEDSTDVDEEFQKEFSSTFPRSEIPNLEKECKYLRTFNSVDELFDFILFDDNGIA
                     KLSDDVEIHFQEDQSLFKVLQFEETGKGNKVQTLVATIPSKILFPDMINSISTINDNK
                     IFDPPTFGITIIGSSHGFDPKKSTTGFVLWINKRGIMVDPPLNSSSFLQSQGVPTRMV
                     DHIILTHCHADHDSGTFQKLLEEYQITVVTTPTILGSFLRKYGALSNLSTDLLRRLFI
                     FRPVMIGEPMIISGAEFRFFYTIHTIPTISFEVFYGGKSIFYSGDTCYDPNRIKDMNK
                     RGIMKTSRMKFFLRPWNHTVVLHEAGVPPIHTPVSVLRALPDEVKNRLYLVHISEHTL
                     PAGSGLKIAKEGVAHTLSLDVMKSSHSEAVDILKLVESVDIFRSIPLTQACEILQTAT
                     KRKYSQGSVIIARDTEPDAFYVVASGVVCVNIGELKKNLIVGDYFGEMSLVMGGLRSA
                     NVQAVTDVEVLSFNKEDFLSITRNSTESIQFITRLWEMRNEKSWETMSLNSVFSRCTN
                     SQKTAIQSILVRELIKKDETLWCKGEEALFGCLVAEGSFVFKEDDSLESFSQGSFLGD
                     INAMTTQPPSIHKTTVVAKEESVIYKVLSQDLIKFFSNNPGIQLAFLDTIFVDALRDQ
                     VQQLVNSKLTY"
CDS                  complement(21566. .21865)
                     /note="GeneID exon scores (in order of location ranges):
                     24.15 - GSCJ_ID dd_00638"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO51112.1"
                     /db_xref="GI:28828504"
                     /translation="MTQLGYQQIPPGMHNNNNNNPNNNSNSSNNNNNSNFIDFLRQSG
                     ITMNDDVNSSGMMTPSSLLNDSSLEHRVASTNKELKDLEDKLQKSSPLNTQSYYR"
CDS                  complement(join(21982. .25164,25267. .25396,25496. .25821,
                     25923. .26082,26307. .26464,26602. .26691))
                     /note="GeneID exon scores (in order of location ranges):
                     340.18, 15.33, 31.36, 13.30, 15.31, 10.95 - GSCJ_ID
                     dd_00637"
                     /codon_start=1

  Query Match             4.2%;  Score 56.8;  DB 2;  Length 266544;
  Best Local Similarity   42.8%;  Pred. No. 0.69;
  Matches  339;  Conservative    0;  Mismatches  452;  Indels    1;  Gaps    1;

Qy          1 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 60
              || | |  |||||||||  ||  |   | |   ||      |||  |  | ||||| ||  |
Db     151228 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 151169

Qy         61 AAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCT 120
              |  || ||   | |  | |    | |  | ||| || | ||| ||| |      |||| |
Db     151168 ATTATTATTATTATTATTATTATTATTTTATTATTATTATTATTATTATTATTATAT 151109

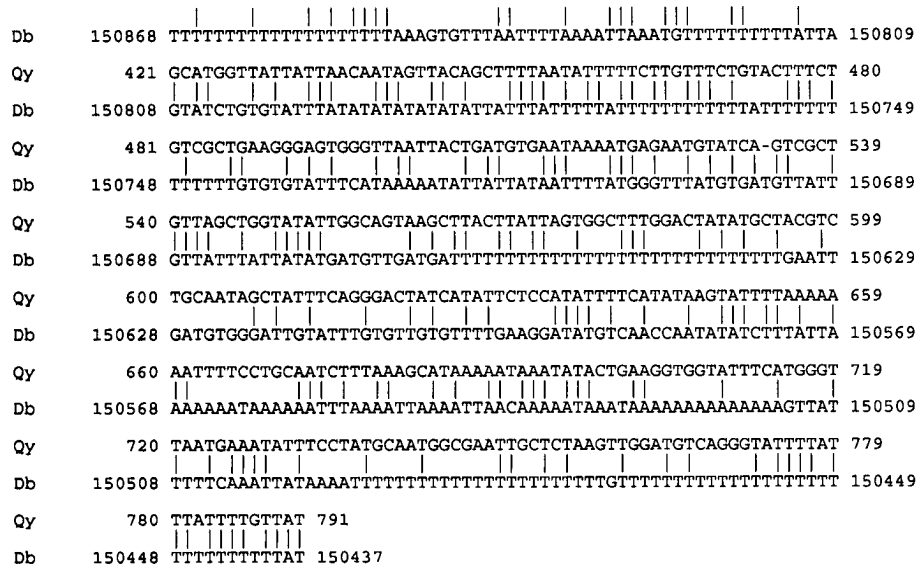
Qy        121 AAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTT 180
              ||      |    ||| | |      | |||   |  |     |   || || ||
Db     151108 TAATTATTATTATTATATCAATTATGGAATTATTTTATTATTATTATTATTATTTTA 151049

Qy        181 GAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAA 240
              | |     ||| |    || |  |       ||||    |   |          | ||
Db     151048 TTACTATTTTTATAATTATTATTATTAATTACTGCTATTATTTTTAATAATCTTATTAAT 150989

Qy        241 TATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCG 300
              |  ||| | | ||    || |  |   | |   ||    | |   |  |    |||
Db     150988 ATTATTTAATTTAAAAAGTATTTATTTATTTATTTTATTTGTTATTTATTTATTTTATTT 150929

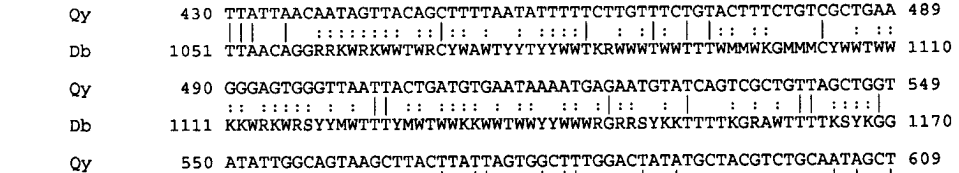
Qy        301 TTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTC 360
              ||  ||   ||    ||  | |   ||  |  |   |  |||||| || ||||| || |
Db     150928 TTTATTTATTTATTTATTTATTTAATATACAATATGGGTTGCTACTATTATTATTATTT 150869

Qy        361 GGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGG 420
```



```
RESULT 7
AY701231
LOCUS       AY701231               1522 bp    DNA     linear   INV 15-FEB-2005
DEFINITION  Orconectes forceps cytochrome oxidase subunit I-like (COI) gene,
            partial sequence; mitochondrial.
ACCESSION   AY701231
VERSION     AY701231.1  GI:58917897
KEYWORDS    .
SOURCE      mitochondrion Orconectes forceps
  ORGANISM  Orconectes forceps
            Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
            Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
            Astacoidea; Cambaridae; Orconectes; Procericambarus.
REFERENCE   1  (bases 1 to 1522)
  AUTHORS   Taylor,C.A. and Knouft,J.H.
  TITLE     Historical influences on genital morphology among sympatric
            crayfishes: systematics and gonopod evolution in the genus
            Orconectes (Cambaridae)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1522)
  AUTHORS   Taylor,C.A. and Knouft,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2004) CBD, INHS, 607 E. Peabody Dr., Champaign,
            IL 61820, USA
FEATURES             Location/Qualifiers
     source          1. .1522
                     /organism="Orconectes forceps"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:306186"
     gene            <1. .>1522
                     /gene="COI"
     misc_feature    <1. .>1522
                     /gene="COI"
                     /note="similar to cytochrome oxidase subunit I"
ORIGIN

  Query Match             4.2%;  Score 56.2;  DB 2;  Length 1522;
  Best Local Similarity   15.7%;  Pred. No. 7;
  Matches   60;  Conservative  166;  Mismatches  154;  Indels    1;  Gaps    1;
```



```
RESULT 8
AE014829
LOCUS       AE014829               250078 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 10 section 1 of 7 of the
            complete sequence.
ACCESSION   AE014829 AE014185
VERSION     AE014829.1  GI:23494865
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 250078)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
  PUBMED    12368864
REFERENCE   2  (bases 1 to 250078)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .250078
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="10"
     repeat_region   1. .842
                     /rpt_type=tandem
     repeat_region   962. .995
                     /rpt_type=tandem
     repeat_region   1165. .1226
                     /rpt_type=tandem
     repeat_region   1165. .1225
```

```
                     /rpt_type=tandem
repeat_region        2096. .2758
                     /rpt_type=tandem
repeat_region        2261. .2922
                     /rpt_type=tandem
repeat_region        2990. .3567
                     /rpt_type=tandem
repeat_region        9425. .9594
                     /rpt_type=tandem
repeat_region        9426. .9563
                     /rpt_type=tandem
repeat_region        9556. .9594
                     /rpt_type=tandem
repeat_region        9840. .9873
                     /rpt_type=tandem
repeat_region        9880. .9944
                     /rpt_type=tandem
repeat_region        9886. .9964
                     /rpt_type=tandem
repeat_region        10025. .10204
                     /rpt_type=tandem
repeat_region        10079. .10204
                     /rpt_type=tandem
repeat_region        10080. .10213
                     /rpt_type=tandem
repeat_region        10225. .10307
                     /rpt_type=tandem
repeat_region        10301. .10359
                     /rpt_type=tandem
repeat_region        10383. .10483
                     /rpt_type=tandem
repeat_region        10396. .10541
                     /rpt_type=tandem
repeat_region        10693. .10777
                     /rpt_type=tandem
repeat_region        10719. .10795
                     /rpt_type=tandem
repeat_region        10870. .11098
                     /rpt_type=tandem
repeat_region        11196. .11235
                     /rpt_type=tandem
repeat_region        11254. .11347
                     /rpt_type=tandem
repeat_region        11383. .12436
                     /rpt_type=tandem
repeat_region        12702. .17518
                     /rpt_type=tandem
repeat_region        13365. .18358
                     /rpt_type=tandem
repeat_region        16920. .19680
                     /rpt_type=tandem
repeat_region        16935. .20332
                     /rpt_type=tandem
repeat_region        16935. .19680
                     /rpt_type=tandem
repeat_region        19528. .19698
                     /rpt_type=tandem
repeat_region        19703. .19903
                     /rpt_type=tandem
repeat_region        19987. .20331
                     /rpt_type=tandem
repeat_region        19987. .20121
                     /rpt_type=tandem
repeat_region        20045. .20265
                     /rpt_type=tandem
repeat_region        20175. .20310
                     /rpt_type=tandem
repeat_region        20402. .20538
                     /rpt_type=tandem
repeat_region        20446. .20501
                     /rpt_type=tandem
repeat_region        20455. .20610
                     /rpt_type=tandem
repeat_region        20605. .20740
                     /rpt_type=tandem
repeat_region        20622. .20825
                     /rpt_type=tandem
repeat_region        20786. .20955
                     /rpt_type=tandem
repeat_region        20935. .21030
                     /rpt_type=tandem
repeat_region        21037. .21322
                     /rpt_type=tandem
repeat_region        21269. .21467
                     /rpt_type=tandem
repeat_region        21529. .21636
                     /rpt_type=tandem
repeat_region        21590. .21994
                     /rpt_type=tandem
repeat_region        21625. .21998
                     /rpt_type=tandem
repeat_region        21632. .22441
                     /rpt_type=tandem
repeat_region        21632. .22001
                     /rpt_type=tandem
repeat_region        21663. .22441
                     /rpt_type=tandem
repeat_region        22727. .22829
                     /rpt_type=tandem
repeat_region        22762. .22842
                     /rpt_type=tandem
repeat_region        22875. .23014
                     /rpt_type=tandem
repeat_region        22998. .25249
                     /rpt_type=tandem
repeat_region        23670. .23959
                     /rpt_type=tandem
repeat_region        23904. .24125
                     /rpt_type=tandem
repeat_region        23989. .24150
                     /rpt_type=tandem
repeat_region        24430. .25018
                     /rpt_type=tandem
repeat_region        24527. .24663
                     /rpt_type=tandem
repeat_region        24530. .24751
                     /rpt_type=tandem
repeat_region        24617. .24655
                     /rpt_type=tandem
repeat_region        25405. .25446
                     /rpt_type=tandem
repeat_region        25464. .25567
                     /rpt_type=tandem
repeat_region        25572. .25734
                     /rpt_type=tandem
repeat_region        25763. .26236
                     /rpt_type=tandem
repeat_region        26240. .26353
                     /rpt_type=tandem
repeat_region        26382. .26411
                     /rpt_type=tandem
repeat_region        26455. .26494
                     /rpt_type=tandem
repeat_region        26541. .26574
                     /rpt_type=tandem
repeat_region        26638. .26753
                     /rpt_type=tandem
repeat_region        27339. .27375
                     /rpt_type=tandem
repeat_region        28248. .28336
                     /rpt_type=tandem
gene                 <28491. .>36165
                     /locus_tag="PF10_0001"
                     /note="Pf EMP1"
mRNA                 join(<28491. .33893,34843. .>36165)
                     /locus_tag="PF10_0001"
```

```
    CDS             join(28491. .33893,34843. .36165)
                    /locus_tag="PF10_0001"
                    /codon_start=1
                    /product="erythrocyte membrane protein 1 (PfEMP1)"
                    /protein_id="AAN35201.1"
                    /db_xref="GI:23494866"
                    /translation="MAAQSSGGGGGCGEEDKDAKYMFDRIGKEVHDEVKNAANVFKDY
                    LKGNLTISTIFGEETVAFTDPCELIKEKRDELLAARGDPCGKGKEHRFSKERGAECDD
                    NKIEGNVRSKGKGGKSAGACAPYRRLSLCNKNFPNMNSKDSLKAKNDLLVDVCYAAKY
                    EAESLIPYHDQYKLTYGDSQICTVLARSFADIGDIVRGKDLFLGHNQRKKKLEERLEQ
                    MFKNITHSNAAKLSVLSSKEIREYWWALNRDQVWKALTCDEENKLGGNAYFHATCSER
                    NGGCSQAHEKCRCPMTSDGKPNDQVPTYFDYVPQYLRWFEEWAEDFCRKKKIYVDIVK

  Query Match             4.2%;  Score 56.2;  DB 2;  Length 250078;
  Best Local Similarity   43.1%;  Pred. No. 0.9;
  Matches  318;  Conservative    0;  Mismatches  418;  Indels    1;  Gaps    1;

Qy          1 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 60
Db     109008 ATTATTATTATTATTCAATTATTTTATTTATATACATAATATATTTTTTATTTCCTTATT 109067

Qy         61 AAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCT 120
Db     109068 TTTATATTTCTTTTTTATTTAATTTTATTTTTTATTTTTTTTTTTTTCTTAATATATTTA 109127

Qy        121 AAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTT 180
Db     109128 ATATTTTTACATGTATTCTTTTTCTTATTATATTATTTATATATATAATATATTATATAC 109187

Qy        181 GAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAA 240
Db     109188 ATATTATATATATCATAAATATAAAATAATTCAACCTAATATATATATATAAAATCTAAT 109247

Qy        241 TATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCG 300
Db     109248 AAAAATTTATGTATATATTCTATATTCAAATAGAAAAAAAAAAGAATACTTTCAATTTTT 109307

Qy        301 TTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTC 360
Db     109308 TTTTTGTTTTTTCATTTAGTTATGGATAAGAAAACATTTCTTATTAAAAAAATATGTATT 109367

Qy        361 GGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGG 420
Db     109368 ACTATAAAATTTTCAGATTATTTAATAATTATAATGAAAATTATATAATTTAGGTTATTA 109427

Qy        421 GCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCT 480
Db     109428 AAATTTTTAGAAA-AAAACTAGGCATAATTTTTTTTTTTTTTTTTTTTTTATACTTTTT 109486

Qy        481 GTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTG 540
Db     109487 ATTTTTGTACGTGTTATTTTTGGTTTTTAATTAAATAACTTTATTATTAATATTTATAAA 109546

Qy        541 TTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCT 600
Db     109547 TATATTAATAAATATCTGTTTAAATTTTTTTCCCATGACAGATGATTGATTTATAATATA 109606

Qy        601 GCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAA 660
Db     109607 AAAATACATATCTCAGTATGCTTAAAAGAATAGTAAGAATAATAATATATTATTTATTAT 109666

Qy        661 ATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTT 720
Db     109667 ATAATATTATATATTTTAAAATATTATTATATATGTATTATTTTAAATACTTAATATTTT 109726

Qy        721 AATGAAATATTTCCTAT 737
Db     109727 TACATTATTTTTTTTAT 109743


RESULT 9
AR579680
LOCUS       AR579680               1141 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 22 from patent US 6784342.
ACCESSION   AR579680
VERSION     AR579680.1  GI:56583130
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1141)
  AUTHORS   Kunst,L. and Clemens,S.
  TITLE     Regulation of embryonic transcription in plants
  JOURNAL   Patent: US 6784342-A 22 31-AUG-2004;
            The University of British Columbia; Vancouver;
            CAX;
FEATURES             Location/Qualifiers
     source          1. .1141
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             4.2%;  Score 56;  DB 6;  Length 1141;
  Best Local Similarity   11.1%;  Pred. No. 8.5;
  Matches  122;  Conservative  401;  Mismatches  556;  Indels   17;  Gaps    3;

Qy         79 ACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCTAAAGACGCACAGGGCTAT 138
Db         53 MSKSRKWTWARMYCKYRRWYNNKSRWWKGWYKKKWYBCANNTSBRYHARRWKDMKTAYBM 112

Qy        139 TATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTTGAATTGGGGCTATCAACG 198
Db        113 TMTNKWGKTGWRHRYWRWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNA 172

Qy        199 GTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTATTCTGAACGA 258
Db        173 TGWDDDTKYHMWNNNGCBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDD 232

Qy        259 GATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCGGTTGGCAATA 318
Db        233 MCKRKVRRWVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRA 292

Qy        319 AAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGTATGTT 378
Db        293 KYWGWNRABVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNN 352

Qy        379 TTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGGTTATTATTAACA 438
Db        353 NNWTYKKARHBARWDWVWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWA 412

Qy        439 ATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGG 498
Db        413 WYWKMDMDWBGTYNNNNNGGRTYYGWTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTW 472

Qy        499 TTAATTACTGAT-----GTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATAT 553
Db        473 WKMKTYWNNCYWKSMTNGKSHRBAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBV 532

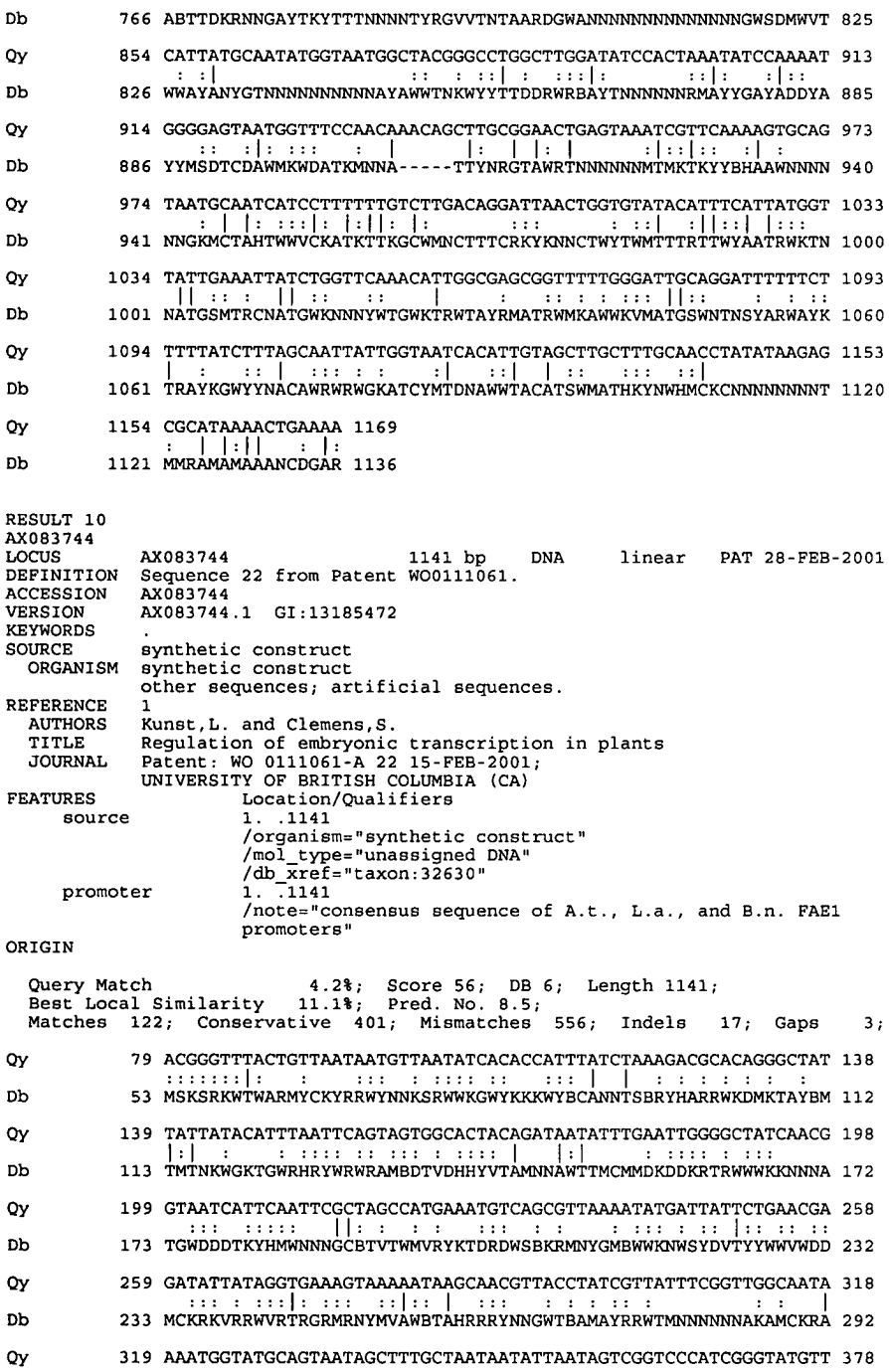
Qy        554 TGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTT 613
Db        533 CSKWWNNYAAWYTKSSWNYTSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTR 592

Qy        614 CAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAAT 673
Db        593 WWBWSHTWBHBRAGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNNNN 652

Qy        674 CTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTC 733
Db        653 NNATCARDDYYAASRWYAMANAKWYY-------YKBAANNAYYTHANNWWGCWNNATDTR 705

Qy        734 CTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTATGA 793
Db        706 RTMWKNNNNNNAGTWKNNNNNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADAA 765

Qy        794 CCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTTTAA 853
```

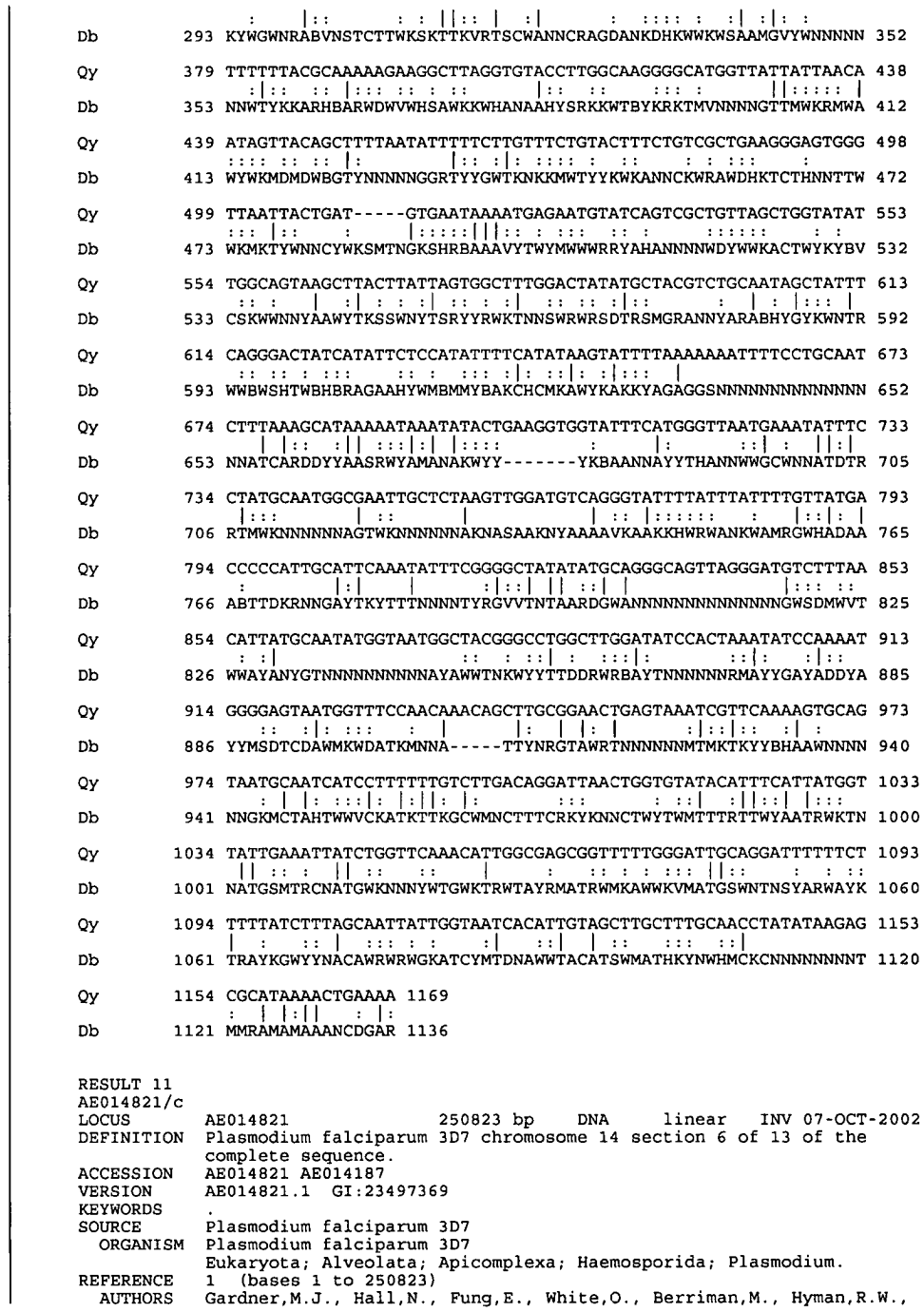


RESULT 10
AX083744
LOCUS AX083744 1141 bp DNA linear PAT 28-FEB-2001
DEFINITION Sequence 22 from Patent WO0111061.
ACCESSION AX083744
VERSION AX083744.1 GI:13185472
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Kunst,L. and Clemens,S.
TITLE Regulation of embryonic transcription in plants
JOURNAL Patent: WO 0111061-A 22 15-FEB-2001;
UNIVERSITY OF BRITISH COLUMBIA (CA)

FEATURES	Location/Qualifiers
source	1. .1141
	/organism="synthetic construct"
	/mol_type="unassigned DNA"
	/db_xref="taxon:32630"
promoter	1. .1141
	/note="consensus sequence of A.t., L.a., and B.n. FAE1 promoters"

ORIGIN

Query Match 4.2%; Score 56; DB 6; Length 1141;
Best Local Similarity 11.1%; Pred. No. 8.5;
Matches 122; Conservative 401; Mismatches 556; Indels 17; Gaps 3;



RESULT 11
AE014821/c
LOCUS AE014821 250823 bp DNA linear INV 07-OCT-2002
DEFINITION Plasmodium falciparum 3D7 chromosome 14 section 6 of 13 of the complete sequence.
ACCESSION AE014821 AE014187
VERSION AE014821.1 GI:23497369
KEYWORDS .
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 250823)
AUTHORS Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,

```
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 250823)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .250823
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="14"
     repeat_region   147. .196
                     /rpt_type=tandem
     repeat_region   235. .301
                     /rpt_type=tandem
     repeat_region   386. .449
                     /rpt_type=tandem
     repeat_region   497. .536
                     /rpt_type=tandem
     repeat_region   715. .775
                     /rpt_type=tandem
     repeat_region   826. .900
                     /rpt_type=tandem
     repeat_region   832. .900
                     /rpt_type=tandem
     repeat_region   949. .975
                     /rpt_type=tandem
     repeat_region   1146. .1179
                     /rpt_type=tandem
     repeat_region   1181. .1224
                     /rpt_type=tandem
     repeat_region   1265. .1302
                     /rpt_type=tandem
     repeat_region   1403. .1546
                     /rpt_type=tandem
     gene            <1574. .>3492
                     /locus_tag="PF14_0301"
     mRNA            join(<1574. .1610,1842. .1912,2027. .2161,2320. .2515,
                     2771. .2873,3001. .3259,3424. .>3492)
                     /locus_tag="PF14_0301"
     CDS             join(1574. .1610,1842. .1912,2027. .2161,2320. .2515,
                     2771. .2873,3001. .3259,3424. .3492)
                     /locus_tag="PF14_0301"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAN36914.1"
                     /db_xref="GI:23497370"
                     /translation="MDFPKLLNKPSLDLLKNDFPLSSKFELEHSSISKHPFLKSGFTF
                     SNNTYSIYTNFKNNIYNSKNEIKFDNSGISLLDIKYEPGFIKNLNLCGKYTKTNGKED
                     DTFEVYSEYTTDDMNIFSSVNVRNFAFKYIHVSSHPKYRNFKFGGLLEGNMDINNLQY
                     SFGGAYTKHHKDNLYIFSLRSIPSNKHFYGSLALNLFFQNKSINDNAISVEVIQNIME
                     KKANINVASIWYLDNKNTFVKTKISNDTKVALSLTHKYNEFVTITLGSQVDISKMSLP
                     DNTKFGMKLYLKS"
     repeat_region   1700. .1763
                     /rpt_type=tandem
     repeat_region   2537. .2576
                     /rpt_type=tandem
     repeat_region   2902. .2973
                     /rpt_type=tandem
     repeat_region   3354. .3388
                     /rpt_type=tandem
     repeat_region   3693. .3717
                     /rpt_type=tandem
     repeat_region   3897. .3935
                     /rpt_type=tandem
     repeat_region   4037. .4102
                     /rpt_type=tandem
     repeat_region   4037. .4099
                     /rpt_type=tandem
     repeat_region   4381. .4443
                     /rpt_type=tandem
     gene            <4492. .>7389
                     /locus_tag="PF14_0302"
     mRNA            <4492. .>7389
                     /locus_tag="PF14_0302"
     CDS             4492. .7389
                     /locus_tag="PF14_0302"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAN36915.1"
                     /db_xref="GI:23497371"
                     /translation="MKEKYVLIKKILNKNNDIISNDTTVSCCDNILCTSSCSSLDEKC
                     DEEEINKYDDFLKGTSRLSSLLKSPYFSSISSYRPCKGLSNNASFDLSSHKSSFDVSS
                     HKSSFDVSSHKSSFDLSSHKSSFDLSSQKSSFDLSSHKSSFSSSSCPPSSTFSYASSV
                     TLSSRKERKKWDIRTHGSLYKKWSCNKNDDKDMWIHRMNFHKHPNHIVKKKKHDYNKE
                     HKSAYCKNGDDVNDEEYLKRMTKNIENYIRKMKDMIYYKRQIKKTKKRKEKKRKEKKI
                     KEKKRKKKIKRSVNIKSICNDNNIKEYDNMNYNYINYKDHLNYENNNINDYNIHHYKK
                     DIQKLYEHLNYEQNNLSTSLLENHYINKNETDHNLFWTCCSEENVEKPREMVFQKNCD
                     LKKKSNEIFISHTQNTFPYNNNLSCICCARIYLIDKIIKVTSNKIKPLCICCFDILLN
                     LIGINTIICISCFKCYPPFEFVNKCVICNSFDMINLKKEDKRNMNIVKKELHIFMELY
                     KYKFSDHFNSDIFLNYFINMNYCFKYLYYYFTIHQQIVISNNFVLARYTFPFSQLLQI
                     IQKDKHKNKMNMEKNMRNEFHDDKHEFDDEEKYKNEIQLKIVNTKENDHVINHMNSHL
                     SDHVNNNIKDNINHNMCVNTQNELHDEYTNKVNEYYQIKNFEGKNKKHIGEQTFEELN
                     SIQNKDAQEEYNIIQEIKEDIMMGERNNKKENYIMKNVETNKEKEQIIKEKEMKHIEN
                     DTIIKDDKVVEKEVNDKINVVNVKDIVDKINVVNVKDIVDEINVVNVKDIVDKINNLD
                     NPNKNDKENNINLNTVHINNKKKEDKKEEICSEKRNKDILIVEKKNKDIFNKKKSINI
                     RNKLYKNKDKGNSKDSVENFFYTFSKTIISHIEKLKNKKIKFSPNEFKNTHIKLNELN
                     FNNTIRKYYANGDLILYYIKYLQRQEKEIHKLYQNYKVIIKKKKKKKKKKK"
     repeat_region   4753. .4920
                     /rpt_type=tandem
     repeat_region   5281. .5336
                     /rpt_type=tandem
     repeat_region   6760. .6876
                     /rpt_type=tandem
     repeat_region   7353. .7386
                     /rpt_type=tandem
     repeat_region   7397. .7455
                     /rpt_type=tandem
     gene            <7677. .>9677
                     /locus_tag="PF14_0303"
     mRNA            <7677. .>9677
                     /locus_tag="PF14_0303"
     CDS             7677. .9677
                     /locus_tag="PF14_0303"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAN36916.1"
                     /db_xref="GI:23497372"
                     /translation="MPKNMEVYKIINEFGELIEEVSLQNNTYEKNNNILINYYLKNNI
                     FRNDKSLLRQKKKRNISINKLNENYELLKENINSVLNILLFLYDTFNKNKNLNQLDLF
                     HILKKIKMFLYDKEIKKLYNHFNDDQYNHNKNNKTTKSNEITKWKEKNVEKCDHILSR
                     NDIRPEIICILNIRKKNMNNFLINNNNKVQIINKNCLQLFNEPYNFDLIYINNIYFKH
                     FISNKIINKYIHNFVNKNNVVLFNFSINYLKNKKKISFFNSFHDYFMNVDKTKINTYI
                     KHQLHKKNEEKKNFINNNEENHQHINEYKINSVQHINKCSQKCMNRKKKKKKKYSDKK
                     GKNILSSKNRKNQEKNIKYKKKTLFKKIINELNNKLNQKYSNEKYNLTLNIYAQKKKK
                     IYNINKCLFRNETNLYYQKNDIHKKEHKVEIMLINNYDNNTNHKINEKIIQKEDMNKC
                     KENEDSFQDRKNNLYDENNIDHTYYDEDKTNPFLPSYGMDIKNIKNINKYINFNNMNK
                     FKKSSIIFLIQLELFLTSDENINIPKKKILHFTLVDINIPILYILKKKNNVNYFNEFK
                     KKQNFDNIFINYIQSYFFNTNVNHYSHNIITHLFQKNPYICFLLYIHDDFIFTNNSEY
                     NKLHQIHNVKKENYSRNMILQLLYIYYMCNICECFNFIKQHC"
     repeat_region   10309. .10336
                     /rpt_type=tandem
     gene            complement(<10401. .>15333)
```

```
                     /locus_tag="PF14_0304"
     mRNA            complement(join(<10401. .12590,12766. .>15333))
                     /locus_tag="PF14_0304"
     CDS             complement(join(10401. .12590,12766. .15333))
                     /locus_tag="PF14_0304"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAN36917.1"
                     /db_xref="GI:23497373"
                     /translation="MEEIEEKYKIIIDALYDTWSNDHDKRKESEKILYEIEKDEKFII
                     CLFDIYTSKSIHYNIRKLGIIYCKNLIVRYWNNRDGFHYSDNTKKIIKKKILDILNNV
                     EYLNNYREFSILLKRISRYELVHNYPELLDCLLYNINIHKTNINNIYIYIYLLYKILR
                     EQYSKKLFKDKKETFDISEKFIKSLEFFWNNSINLYFQNNATDIFCMCNDEKNCNKCK
                     NLNSLINLYEKQLFVSNVDDTIRNEGTANENIPEDENYYIHDKKIHMMKYLDSIILNL
                     IINRNDIKKVNNYVNSKMSKENKGDQENKVDQEYKVDQEYKVDAEYKVDQEYKVDAEY
                     KVDPEFKVDQEYKVDAEYKVDSEYKECKENYKNKESPNSSNNTTTTNNNNNNNNNNNN
                     NVNVNVNDTDLNIRTNFLNAFINKIIYYLKFINNNSASYYLKYLKFLLTSFFLMIEFH
                     TSFHVYLKKDITEKLICLFLKKNLSLNNYTNDYEHKFIEIVNLCISILKSIFYHFCLN
                     QYNLIKQKKVRENYINVKNIMCNKNINIENYNVNTLESKIINNFMINVKSMENENNIL
                     FNKNLSLKDILTYIREEVKQALSNDEFILIHNEKALEIFYFLRIYCINISSEQIIEVT
                     LNIRDDYETEENIFYNNAKDCIIELSSEPFLFSIFHHFFEPLINSFHNYVHFLKGLPV
                     HLKNIQIPEFSSAIIQLDGYLNIYYILYPSFHKKIKIEHILCMIRFFMDFLNIKYTHP
                     LISYRIALIIKIWTKCYKTSFPFIEDLILLIYENIRFLYIQLLQKNVVSLITENIPSY
                     VNDMTFQHYAIKSVNFLPILFFKFISLFKYFFKYEYVYKNYDYINEFLVSCLINVLTK
                     ITYPKNIQKILKILSHIVFMSNKEKDITIFKDNYNFLLDLYMNSNISIKEYILNILVQ
                     ILNKNYEYVGSLVEYMNKENYHNDNNKVGVNEKVCDDIVLFYFSYDIILYTLSAKIEK
                     QPFKDNIIFNHPHIPSEYKANHNINNDITSDINSDVNNNINNFQFNNIYNNLINFPYI
                     KEINLDIDKTISDNFYSLWICILKLIYKMFHQKKEEIVKKICSFYIATINYLYEYFKN
                     QTLGKISYELSNLCLDVVIEYMCLFIIHETHNFYLTYESINIKMLTNDIIKSTILKVV
                     EFNFILNDEIKIEKCIYILHLCLSIYKNDMKNEKSIIYTYIAQFVIIYLLKVLLKYFS
                     KYFNFYNYIQQNQNKRAERNDANKNYNLHSNNNNNFNDDVQKNGNVIFDEYKTIGEEK
                     SNRNKNNIPDISFDLQGNNTNNYYLNHFKNDVNNNDNDDNNNNKNNDNNNDDNNNNNN
                     NNNNNNNSSSNSEQINNKLNQCNKNHIKYIYDNIINYQHDENFLNEFTHFQNDVKKLF
                     NINYINDYDITKTFLNFNFSTINEDNYSYINKHSVFTLISILSLYEPNYLNYILITFL
                     TYFKINIPIFIFSFFYQSQYIHNKNILNSLLIFIYILTQNYILTPFPSIFLQTNFPSI
                     TNTQSFEKLMYDTNKNNHLFTCNKDEHTLLIIEILKLINNILNTNKDIYIEKKNILHL
                     STISTNLNGTKIFHSYNYNYILRKVLKYDDFAS"
     repeat_region   11177. .11276
                     /rpt_type=tandem
     repeat_region   12606. .12658
                     /rpt_type=tandem
     repeat_region   14155. .14194
                     /rpt_type=tandem

  Query Match             4.1%;  Score 55.6;  DB 2;  Length 250823;
  Best Local Similarity   49.0%;  Pred. No. 1.1;
  Matches  148;  Conservative    0;  Mismatches  154;  Indels    0;  Gaps    0;

Qy       428 TATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTG 487
             |||| || | | || |||    ||||| |||| || || ||| | || |||       |
Db    173394 TATTTTTTATATTATTTATTTATTTTATTATTATTATTTTTTTTTTATTTTGGAATTTTC 173335

Qy       488 AAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTG 547
               |   |       || ||| ||    || ||||  | ||   |  | |         |
Db    173334 GACCCATAATAGAAAATACAGAGAACAAAAAAAAAAAAAAAAAAAAAATGTTGCCATAATA 173275

Qy       548 GTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATAG 607
               | |||    |  |   ||| ||| |  |   |||    | ||| |        ||||
Db    173274 TTTTATAACTACAACCGTTATTTAATTTTTTTTTTTTTTTTTATACATATATATAAATAA 173215

Qy       608 CTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCC 667
             ||| |     |     ||||      || | | ||||| || ||||||   |  | |
Db    173214 ATATATATTATATATATATATGTATAGTAATATTATATATGTTTTTTAATTTAGGTCTAG 173155

Qy       668 TGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAA 727
               |  | |||| || |||||| ||| | |   ||     ||| | ||   |  || ||
Db    173154 AGTATGCATTAATGCTTAAAAAAAAAAAAAAAAAAAAAAAAATATATTATATATATAAAT 173095

Qy       728 TA 729
             ||
Db    173094 TA 173093
```

```
RESULT 12
AC143489/c
LOCUS       AC143489               83313 bp    DNA     linear   HTG 09-APR-2003
DEFINITION  Macaca mulatta clone CH250-267D17, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC143489
VERSION     AC143489.1  GI:29648769
KEYWORDS    HTG; HTGS_PHASE2; HTGS_PGI.
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Cercopithecidae; Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 83313)
  AUTHORS   Csuros,M. and Milosavljevic,A.
  TITLE     Pooled genomic indexing (PGI): mathematical analysis and experiment
            design
  JOURNAL   (in) Guigo,R. and Gusfield,D. (Eds.);
            ALGORITHMS IN BIOINFORMATICS, SECOND INTERNATIONAL WORKSHOP, WABI
            2002, ROME, ITALY, SEPTEMBER 17-21, 2002, PROCEEDINGS: 10-28;
            Springer (2002)
REFERENCE   2  (bases 1 to 83313)
  AUTHORS   Milosavljevic,A., Sodergren,E., Csuros,M., Li,B., Jackson,A.R.,
            Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L.,
            Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
            Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Egan,A., Earnhart,C., Edwards,C.C.,
            Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
            Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
            Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
            Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
            Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
            Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F.,
            Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
            Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,
            Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
            Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
            Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
            Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
            Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
            Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B.,
            Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
            Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R.,
            Washington,C., Watlington,S., Williams,G., Williamson,A.,
            Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
            Zorrilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 83313)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
```

```
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
            Center project name: LBOF
            Center clone name: CH250-267D17
          -------------- Summary Statistics
            Chemistry: Dye-primer Bodipy: inf% of reads
            Chemistry: Dye-terminator Big Dye: inf% of reads
            Consensus quality: 9313 bases at least Q40
            Consensus quality: 10835 bases at least Q30
            Consensus quality: 12380 bases at least Q20
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
          * NOTE: The contigs are based on the application
          * of the PGI method using the Human genome (NCBI build 31)
          * as the comparative genome.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 1 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *       1    83313: contig of 83313 bp in length.
FEATURES             Location/Qualifiers
     source          1. .83313
                     /organism="Macaca mulatta"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9544"
                     /clone="CH250-267D17"
     misc_feature    1. .83313
                     /note="assembly_name:CH250-267D17.1B
                     CONFIDENCE:_0.67"
ORIGIN

  Query Match             4.1%;  Score 55;  DB 14;  Length 83313;
  Best Local Similarity   37.3%;  Pred. No. 2.2;
  Matches  163;  Conservative    0;  Mismatches  274;  Indels    0;  Gaps    0;

Qy        296 TATCGTTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAA 355
              | |  ||| ||   ||   ||| ||||  |     ||| |   |    | | | |  ||
Db      40540 TNTTTTTAATTTTATTTTNAATTAAATTNNAAAATATAAAANNATNTNAATTTAAAAAAA 40481

Qy        356 TAGTCGGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGC 415
                |    |          ||  || ||||     || |       |  | |
Db      40480 AAAATATTANTNAATAAAATAATTNTTTTNAAATAANAANNNTTAAANATTTTAAAAAAA 40421

Qy        416 AAGGGGCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTAC 475
              |        |  | |||  | | ||   ||   |   ||||   |||||         |
Db      40420 ATTNATNTTTNTNATTTNTTANAACCATTNNANAATTTANNNTTTTTTAANNNNTTTNNA 40361

Qy        476 TTTCTGTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGT 535
              ||| |       ||   | |   ||  | | | |    ||| |||  | ||   || |
Db      40360 TTTTTNANTNNAAATTTAATAATTTTTTNAATAANNAAAATTAAAANAAAANANATTATA 40301

Qy        536 CGCTGTTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTA 595
                | ||  |  |   ||    ||       ||||      |   |  ||||   |
Db      40300 NANTNATAATTTNTTANTTTNTTTTANNAANTAAAATTAAAATATNNTTAANATATNTAA 40241

Qy        596 CGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTA 655
                    |    ||| |   |   | || |    | | |   |  | |  |||
Db      40240 TTNNAATTTTNAAANTTTNATNAAANTTANTANTTANTAAAATNATANNTTAANATTAAN 40181

Qy        656 AAAAAATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCAT 715
              | || | |   |    | ||  ||    ||| || | |   | |       | || |
Db      40180 TANAATTNTNANTTTNNTATTANAAAANNNAAATTANAAANTNTAANTNNTNAAATTNAC 40121

Qy        716 GGGTTAATGAAATATTT 732
                ||   |  ||| |
```

```
Db      40120 AANATATATATNTATNT 40104


RESULT 13
AC117076
LOCUS       AC117076              146570 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 3323568-3470138 strain
            AX4, complete sequence.
ACCESSION   AC117076 AC115610
VERSION     AC117076.3  GI:42733926
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 146570)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 146570)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 146570)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 146570)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 146570)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Feb 21, 2004 this sequence version replaced
            gi:28828883, gi:28829496.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .146570
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="3323568-3470138"
     CDS             complement(join(<63. .86,436. .603,692. .704))
                     /note="GeneID exon scores (in order of location ranges):
                     1.46, 2.66, -0.45 - GSCJ_ID dd_03180"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38882.1"
                     /db_xref="GI:42733988"
                     /translation="MTILGLISSISNVKSISKSNNLSSLSNSLSSLQSMNSIERRGGI
                     SISNPGFLGGIGRNPGESMVCDSG"
     CDS             complement(join(1507. .1763,1872. .1884))
                     /note="GeneID exon scores (in order of location ranges):
                     19.50, 3.22 - GSCJ_ID dd_01276"
```

```
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38835.1"
                     /db_xref="GI:42733927"
                     /translation="MTILGSISSIGNVKSITKSNNLSSLSNSSSSTQSMNSIQCGGGC
                     GNGGLLGGVGGLVGGVLVGTGVIIGSVLHGVGSILTGGSNNCGCN"
     CDS             join(2611. .2623,2691. .2947)
                     /note="GeneID exon scores (in order of location ranges):
                     3.18, 23.32 - GSCJ_ID dd_01275"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38836.1"
                     /db_xref="GI:42733928"
                     /translation="MTILGSISSIGNVKLISKSNNLSSLSNSSSSLQSMNSIQCGGGC
                     GNGGLLGGVGGLVGGVLVGTGVIVGSVLHGVGSILTGGSNNCGCN"
     CDS             complement(join(3484. .3740,3824. .3836))
                     /note="GeneID exon scores (in order of location ranges):
                     25.27, 0.79 - GSCJ_ID dd_01274"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38837.1"
                     /db_xref="GI:42733929"
                     /translation="MTILGSISSIGNVKSISKSNNLSSLSNSSSSLQSMNSIQCGGGC
                     GNGGLLGAVGGLVGGVLTGTGVIVGSVLHGVGSILTGGPNNCGCN"
     CDS             join(4219. .4228,4659. .4915)
                     /note="GeneID exon scores (in order of location ranges):
                     -7.18, 24.38 - GSCJ_ID dd_03039"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38838.1"
                     /db_xref="GI:42733930"
                     /translation="MVKCSISSIGNVKSISKSNNLSSLSNSSSSLQSMNSIQCGGGCG
                     NGGLLGGVGGLVGGVLVGTGVIVGSVLHGVGSILTGGSNNCGCN"
     CDS             complement(join(5405. .5655,5746. .5758))
                     /note="GeneID exon scores (in order of location ranges):
                     25.79, -3.80 - GSCJ_ID dd_03235"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38839.1"
                     /db_xref="GI:42733931"
                     /translation="MTILASISSIGNVKSISKSNNVSLSNSSSSLQSMNSIQCGGCGN
                     GGLLGAVGGLVGGVLTGTGVIVGSVLHGVGSILTGGSNNCGCN"
     CDS             join(6487. .6499,6572. .6825)
                     /note="GeneID exon scores (in order of location ranges):
                     3.30, 26.87 - GSCJ_ID dd_03141"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38840.1"
                     /db_xref="GI:42733932"
                     /translation="MTILASISSIGNVKSISKSNNLSSLSNSSSSLQSMNSIQCGGCG
                     NGGLLGAVGGLVGGVLTGTGVIVGSVLHGVGSILTGGSNNCGCN"
     CDS             join(7061. .7150,7227. .8279,8516. .9178)
                     /note="GeneID exon scores (in order of location ranges):
                     -0.28, 67.69, 40.02 - GSCJ_ID dd_03341"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). MkpA protein"
                     /protein_id="AAS38841.1"
                     /db_xref="GI:42733933"
                     /translation="MDTTKNNDEFDISYFLTCSIFGFILWILTEQILKYYNKTNKNNK
                     YNLPKGPSFLKWFINYLFNFYDLKLSNNKEEDNNNNNNKSNNSLSQEELIEDTSENTV
                     LKWFNQLNSDNYSVSFFGRPMIFTRDTTISKYILSSNNIDNYTKPPDSSGVLIRLAQN
                     SILMSEGDQWRYHRSIINQPFSSKNVKLMIPTIITTINKLINHLNNNNNNNNNNNNNN
                     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTIIIDIHSYCTKLTFDII
                     GKLSIGYDFNSIESSDNDNDNNDDDDISKQFDFILNEMIRPIRRFSSYLPLYNDIKLF
                     KFLNELESIIKGAINSRSLITDNNNNKTYKKNFLLDNLLDDNVKEKDIIGNINTFLLA
                     GHETSANLLTFIFYLLSTHNNVQNDLYNHLIENQKKKINKDNKFTEEDEDYQSIEFLD
                     WVIYETLRLFPPAPMIGRTSKNDDILKSGNNNNNNNNNISIPSETLILISVYAIHRDP
                     KLWKDPNIFNPYRWKNIENINNRSDFIPFSSGGRVCVGQKFSIVEARIIISKLILNFE
                     LSFNNLKSKPFKIYQRATLTPKYPVFLNFKKRENK"
     CDS             complement(9415. .11361)
                     /note="GeneID exon scores (in order of location ranges):
                     133.26 - GSCJ_ID dd_03256"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). CIGB protein (Fragment)"
                     /protein_id="AAS38842.1"
                     /db_xref="GI:42733934"
                     /translation="MMEPINYKNISFNDIKLDSNCNSSSSSSSSSIGSNRSGSGGGGA
                     ISRKSFDCSTSNGISEFGGDDGGCGGGCGGDMGNNINISINKRHNSVNSISSNKNINC
                     LYHDLNDLITYCNDCERMICVQCIVSNDHCGHKFTEISKPFASSLLEKFQKQQFPNLL
                     GYRESDKHIISALDYCCNDVERQHIESVKQVKQHFQKLIESIQSIEKDIIEKLQSTVE
                     WNKELQNSIKQQFERDSKKIDSIISKHHSKINGYQIDQIYDFETKQHIEIIKDSDQCT
                     PLLKRRLEILNNNSSSKIKFKNDLLTFNMEIFDEIKEKVLNFTKIENVEYKDLKSTKQ
                     HLMLNKKIQPYNIDYYLCHGEPIPLGETSVALDSKYSSKSICIQSNKSIDRLLLCNGF
                     KLSLSEFKIPDSIKTLRIGDIKYSWDANSIPSSVTELHLCDGFNHILSSGMIPNTVTK
                     LYLYDIKQPLTMGSIPSSVNELHMCDGFSLQLAPFVLPNSITSLSIGEIKYSILPEFI
                     PSSVKKIRLYSSFNHPIIPQNISESVETLVLFNIKLNLENKVIPKSIKELHLCDGFEQ
                     QIQPFSIENGGKLEKLVLCDIKEKLLKNSIPNSVKKLIFLGYNHQLENGILNDSLELI
                     YFSKSFTHPIDNNLIPNNIEIIKK"
     CDS             complement(join(12693. .12990,13133. .13227))
                     /note="GeneID exon scores (in order of location ranges):
                     18.88, 1.20 - GSCJ_ID dd_02598"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52031.2"
                     /db_xref="GI:42733935"
                     /translation="MSLNNFSPEEKQNIVRSIFAEEEIVYNVFFERGDDSFKNLKLSE
                     IKAILSILKSSRSMAYPLSGKKQVLVDLLTSIGFVGPINTSPQPQPPQPPPSPTPPGL
                     FQEELQPHSFLELDLGITVELDNLLGEN"
     CDS             join(13881. .13926,14100. .14230,14342. .14589,14703. .15129)
                     /note="GeneID exon scores (in order of location ranges):
                     6.31, 2.00, 17.53, 44.25 - GSCJ_ID dd_02599"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52033.2"
                     /db_xref="GI:42733936"
                     /translation="MTLVEYSETLKKSLEEFCKQDKKVVTPELNSIIENISKTGITCY
                     PWNILKELFYFKLNEILDIFEKEYIATSSTTTTSTTTTTTNTSSSSTNSPQSPQSPQS
                     PQLNVNIKDEEMQKIKNQVNLSGLIKMKKEFLDSFKESKLSPFTVQRLCELIINYKMY
                     TSFSKYLCAVEKMLNVSTLPHLTPEEVIEFNKNQNSGSRLNSSTNTTTTTTTTTTTTT
                     PTEKGNNDQPNLDSFAFSSSFTTSFPSDTTSSTSTSLSTDQEMKDLDVEQQKDDTDNT
                     ANENESK"
     CDS             complement(15517. .15972)
                     /note="GeneID exon scores (in order of location ranges):
                     30.20 - GSCJ_ID dd_02600"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52034.2"
                     /db_xref="GI:42733937"
                     /translation="MDPGRSHIETIIESSTHKFILFDEPTEETIPYFKDLMKKNSCIN
                     IVRCCEINYDASLFEGVKIHELCFKDGNVPPKDIIERWLEILKQAFIENGKQKTTVGI
                     HCIAGLGRTPLLVCIALIEDGMKPLQAVEFVRSKRKNAINSPQIKFLKE"
     CDS             complement(join(17629. .18615,18760. .18992,19147. .19282))
                     /note="GeneID exon scores (in order of location ranges):
                     132.77, 35.41, 19.03 - GSCJ_ID dd_02602"
                     /codon_start=1
                     /product="similar to seryl-trna synthetase, cytoplasmic
                     (EC 6.1.1.11) [Schizosaccharomyces pombe]"
                     /protein_id="AAS38843.1"
                     /db_xref="GI:42733938"
                     /translation="MGLDIVLFRADKGGNPDLIRQSQKVRYANVDAVQEVIDLDKVVK
```

```
  Query Match             4.1%;  Score 54.4;  DB 2;  Length 146570;
  Best Local Similarity   44.9%;  Pred. No. 2.2;
  Matches  248;  Conservative    0;  Mismatches  301;  Indels    3;  Gaps    1;

Qy        235 TTAAAATATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTAC 294
              ||||||  |  |||||||      |||| || |   ||||    || |  |  ||  | |
Db      70927 TTAAAAGTTATTTATTCTCCTGAAGATTTTGTTATTGAAGAAGAAGAAGAATAAATTAAT 70986

Qy        295 CTATCGTTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTA 354
              | |   | |    |  || |  |||   | |    |  |   |||  ||    |  ||
Db      70987 TTTTTAATTTCAAAGAGGGAATATAAAACTTTTTTTTTTTTTTTTTTTTTTACATTTGATA 71046

Qy        355 ATAGTCGGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGG 414
```

```
Db      71047 TTTATAGCGCATTCTTATCAAAAATAAATAAAATAAAATAAAAATAAATAAATTATTGTA 71106

Qy        415 CAAGGGGCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTA 474

Db      71107 ATAATAGTAATAATAATAATAATAATAATAATAATAATAATTATAATTATTATTATTATA 71166

Qy        475 CTTTCTGTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAG 534

Db      71167 TGTACGTGGGTTATATTTATTTATTTATTTATTTATTTATTTTAATTATTTATTTATTTA 71226

Qy        535 TCGCTGTTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCT 594

Db      71227 TTTATTTATTTATTTATTTATTTATTCAAATATTTATTAAAAAAAAAAAAAAAAAAAACA 71286

Qy        595 ACGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTT 654

Db      71287 ACTTATTGATTTGTTA---AAGTTTTTTTAATTATTTTTTTTTTTTTTTATATCTTTGTT 71343

Qy        655 AAAAAAATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCA 714

Db      71344 GGATTATTTTTGCCTAAAAAAAAAAAATTAAAAAATAAATAAAAAAAGATTTTAATATTA 71403

Qy        715 TGGGTTAATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTAT 774

Db      71404 AAAACAAATTAATCTTTTTCACTCCAATATCCAATTTTTTTAAAATAAATGTCCTTTTTT 71463

Qy        775 TTTATTTATTTT 786

Db      71464 TTTTAAAATTTT 71475
```

RESULT 14
AC116979
LOCUS AC116979 331039 bp DNA linear INV 21-FEB-2004
DEFINITION Dictyostelium discoideum chromosome 2 map 6445720-6776760 strain AX4, complete sequence.
ACCESSION AC116979
VERSION AC116979.3 GI:42733765
KEYWORDS HTG.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 331039)
AUTHORS Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P., Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K., Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and Noegel,A.A.
TITLE Sequence and analysis of chromosome 2 of Dictyostelium discoideum
JOURNAL Nature 418 (6893), 79-85 (2002)
PUBMED 12097910
REMARK The Dictyostelium Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 331039)
AUTHORS Baumgart,C.
TITLE Direct Submission
JOURNAL Submitted (05-APR-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE 3 (bases 1 to 331039)
AUTHORS Baumgart,C.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE 4 (bases 1 to 331039)
AUTHORS Baumgart,C.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE 5 (bases 1 to 331039)
AUTHORS Baumgart,C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT On Feb 21, 2004 this sequence version replaced gi:28829703.
CDS predictions from GeneID do not necessarily reflect true genes.
Further Information is available from IMB Jena, Department of Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)
and the Univerity Colonge, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shtml
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).

FEATURES Location/Qualifiers
source 1. .331039
/organism="Dictyostelium discoideum"
/mol_type="genomic DNA"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="6445720-6776760"
CDS complement(join(912. .1165,1280. .1427))
/note="GeneID exon scores (in order of location ranges): 30.46, 12.19 - GSCJ_ID dd_02783"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAS38697.1"
/db_xref="GI:42733766"
/translation="MSEKTFTRQEVAKHCSLNSLWIIYNDDVFDVTNFVVEHPGGEEV
LKGNGGKDATQEFDDVGHSASAIAKMQSLRIGRIEGAKPREEKKKEIKKTTTTTSAPK
QQESAGLGLLKIPLIIIVLAIAAYFFMGDQQ"
CDS complement(join(2345. .3098,3189. .3526))
/note="GeneID exon scores (in order of location ranges): 98.40, 31.82 - GSCJ_ID dd_02784"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime mold). Protein disulfide isomerase precursor"
/protein_id="AAO52220.2"
/db_xref="GI:42733767"
/translation="MKILLFVTLIALAFVALCSAEGNVVVLSPDNFDTVVDGSKTVFV
KFYAPWCGHCKKLAPDFEILADTFAPVSNKVVIAKVDCDQADNKALCSKYDVSGYPTL
KIFDKSTTAKDYNGARSVDELLTYINNHAKTNVKVKKAPSNVVDLSPSNFDSVVLDKS
KNVLVEFYAPWCGHCKKLMPDYEILGNTYANEKDVVIAKIDCDAADNKAICSKYGVTG
FPTLKWFGKQSKDGEKYEQGRDLDTFINYINKQAGVNRVKGGKLAVGAGRVEQLDTIA
TEFIAAAAEVRKELVKKAQTVVDSLPEELRTEGSYYVKVMKTIAEKSIDFVTTEIARI
TKLVSGSMSGKKADEFAKKLNILESFKSK"
CDS join(5795. .6179,6292. .6567,6659. .6881,7332. .7363, 7523. .8215,8284. .10451)
/note="GeneID exon scores (in order of location ranges): 14.11, 5.65, 10.80, 0.73, 45.28, 141.12 - GSCJ_ID dd_02785"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime mold). Hypothetical 127.0 kDa protein"
/protein_id="AAO52221.2"
/db_xref="GI:42733768"
/translation="MNINLLKFFIIYYFCFVLINFCNCQTLKITDISNPKNYKVYPQY
SDREFCDFSPEILCYVEPAMDEGDQLMFTTDFDYSEMIRSDRTGSIIGITHTALSFGS
YTSNIVCYLLNNPTIKVDLNISYSCEMIDYELMKVKILKPNYFKKNNLYGNGIIQLIG
LKYPLNNLVIFGGSAYPLNPNYFLISLEELDGGTVENYNISLLFYGPHTIVVTIPAST
YLDYNLDISNSKINYYPNETTNVSEFGLGCSTVYSVTINKTSNDYEPYLYVDDGNSKN
ADFTPIYQDENFITYLGQFNSFVTVTSPYAQILNSSILSTLLEFECKLLFDESYLFRF
KVNTNGYGASIYLGSDITLRYESVVDSESNSSVLEFETVVDLYKDSITKFTISNLLSN
RINYLRGEYYSISPLLKFEIPFGELNPVLITNISFLYNNIDVTNKSVNNIIYFSYEGE
FDKGSSFALLLNDPVSILKNGDLDSSDFRFSKWNSSTSMFQIEFTVPANTQPGHIPFK
LYGNSNPNYNPIDSTMLPLSAQLFIISSNFDSYGPIFNKVEKINTTNEFGWKFVIGDK
INGFDYGDIIVRGEMDSSTYKFHLTPLNLTSGDKYNGEYQINIKISILCASQNYIITD
VYLFDTRGNFAKFSVSEQCESIRNPFINFLGDSTINKIYKKCTTENDGYDISPPELTL
FSAVKTILNDSQTIVIAFDFQAVDLDTGLKNEQYPIVYIESTELQIIQCESKIKSINS
KTANYTCEIELPVGFGYNSDIVFSVYGFINNGGYYSGFPSTYISSAFPIYSMSNIQLV
RKIEILRTSQISSTGGELWIVGRGFKTSIQYVQIKYSTDLVFSQISFPSIIFSTALLI
LDIDPTDKPFTIRVVTKNGQQSNEFIVTPLFSFLSVSPLNTPTPTTTTQLPTPTNKPQ
TCLGEPICGGPKQGFCSSKGCVCYQPWIGNDCNAQVVIIPQPSTNTSQPSTEIPIIDN
NNNQQPSGINYLYKSLISIVSIRELDFNGNQINLYPFDRWIFTQINQNKSQYFTSIKN
NSTTPFNTNITVILEWFKDTSTIEFANSSITMYPSSIKYTIEITEYKFSNQLNNLQLV
MSASFESSKTKEICSLKDFGETSNGDNSNYFKIQINDHSLYGRFIKRAIIDSKISTIE
NVLLDSKMNSIQTASSSQSYIGITIPNYQKSIIIDPDFSVLIDSKSTSKNDENSVCTP
NQSKLTTPQLIGIIIGSIAFAAVIAISIAYHIIKKKKDTKFKNGLQNKLKTFKS"

```
CDS             join(11030. .11042,11114. .11283)
                /note="GeneID exon scores (in order of location ranges):
                -1.76, 12.26 - GSCJ_ID dd_02786"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO52222.2"
                /db_xref="GI:42733769"
                /translation="MTILKSISSMGISNRKNSFSFVSTNSMIETNQNVNSVSKSGYSK
                LIKGAALMAEGISGFF"
CDS             complement(11515. .12534)
                /note="GeneID exon scores (in order of location ranges):
                87.29 - GSCJ_ID dd_02787"
                /codon_start=1
                /product="similar to Burkholderia sp. DBT1. Putative
                oxidoreductase"
                /protein_id="AAM34303.3"
                /db_xref="GI:42733770"
                /translation="MLPITNKVALITKKGFSESFSIVDNEIPKPKGNEVLIKVDYVGL
                NRCDLYFSLGLYLPFENYPNVSGFEASGTVVEFANDELENSSKFKVGDKVSVIYTCDH
                RKYGTFAEYAVFPIGCLVKNNELSSQKEIAANWISFFTAYYGLVEFAQIKKGDYVIIT
                AGSAGTSLAGIAIAKFFGAHTIVTSRNEKKKQQILDYGADYFVSLDTEDIVKRTFEIT
                NGKGANIIYDSVGGQSAQKLFESSSQFGKIVFYGNMNHSEPTPLPILIGLKNYLTIKV
                LNTNEYLYQPEFNKKAIEFLNSNFHHFKSIVGKEFIGIDSLVDALKYLETGDLFGKVV
                VKVNN"
CDS             complement(join(13231. .14884,15364. .15368))
                /note="GeneID exon scores (in order of location ranges):
                104.49, -7.13 - GSCJ_ID dd_02788"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAM34304.3"
                /db_xref="GI:42733771"
                /translation="MINPNLLKPLFIENFCSSSQSWLDLIKNRKLNFYTITCLTHYYS
                TVKVLYFDIKSIFENDKKQNENQQNEKQQQQQQQQQQQQQQNEKSLNDRKSIFKKSYH
                IIRLLIEMNDIINYKDIRIYLPNQSEERKLLMKIREGEEIDIQSCFELVQSKYLEFES
                KFNELKKNSTNDSLIKLNSARALHVLNDWVVDIRRNSIENSNNLNSFKKYKFENENHH
                DESNNSFKIAMDLMNKNNIKGTLIYFGPSGSHLHGLQDDDHNDENSNDNEKIDWIGIY
                AAPTDDIVSLFPLALKHDSSGFIMQSNEHIGINLNPAPDRPNNQNKKDSFSKGIILYE
                LKYALQLIVNCSYRVLECVFMSSNQVLNNHPFKNKFYITDAWKDFINQVLINDNNNNK
                TFSKLSNSNLIQQYWGLAKGISMKNSTIFKMVKSNLLRKKDDAITIAEEDLVTDDSIS
                KNLFFLNYLLIQCNTILNGIVFNENNTIINSPNNEILLKEKEIPSIFIDKQSKDYQFL
                MSLKHNKLNKLSIIQSLIIDSDSILEEIRIKLDTLKKIQRIDLV"
CDS             join(17515. .17664,18062. .19243)
                /note="GeneID exon scores (in order of location ranges):
                1.18, 93.30 - GSCJ_ID dd_00463"
                /codon_start=1
                /product="similar to mitochondrial carrier protein family;
                protein id: At1g14560.1 [Arabidopsis thaliana]"
                /protein_id="AAO52223.2"
                /db_xref="GI:42733772"
                /translation="MNSSDFKKSFKESTENNSNTYRPSKTLNTYKSLKNGNNIANSLI
                IQLNNNYLVSGSIAGAISRSATAGFERLTIIQQVQGMSQNLSQGYVGCIAAMKEMVKR
                EGFKSIWKGNGANIVKVSPNSGIRFLTYEFCKKHFLDNSSNHPSSSSIENGIDGNGVG
                CGSGSEMKMTVPQTMFSGAMAGLTSTFFTYPLDVVRIRLSLQGSCSNDYAAHRYNGIT
                HSFFKIHKDEGVKGLYKGLGTSIASIVPWVSISFATYEGFKIICKKMILNYQISSSSL
                STTTTTPSSINNNNNNNNKNNNSFIYENELGENGINLTNTSGCSTMASTMPSSLLINS
                VASDENELKKGVNMICDFVCGALSGAVTMTVCYPLDVLRRRMMIQGIGGNKVLYKNGW
                DATKKILSNEGLVAFYHGIIPAYFKVVPTVAISFAVYEICKDLGSNKYQQK"
CDS             join(20299. .20406,20551. .20660,20960. .21569,21708. .21828,
                22984. .24188)
                /note="GeneID exon scores (in order of location ranges):
                9.06, -2.15, 55.50, 6.70, 58.02 - GSCJ_ID dd_01980"
                /pseudo
                /codon_start=1
CDS             join(24256. .24266,24388. .24538,24623. .24750,24902. .27884)
                /note="GeneID exon scores (in order of location ranges):
                -2.62, 6.90, 11.20, 259.32 - GSCJ_ID dd_01978"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). Histidine kinase DhkF"
                /protein_id="AAO52224.2"
                /db_xref="GI:42733773"
                /translation="MILKFIIQQCVFKVAKIFSDNLIQIFMDISSVYSSSTLFFSKLS
                LLMCKVFNSKCAGISVLNKDGLFDTIAFCSSTQELENFSYSLKKISKNITYYDYMEIP
                LKDSKEFYFGDIVLFGDQTLNPNILDTNILTFITNRTSMEFEKKKVSDQLIIAKNLLD
                QSPSCSCLVSRNGKIIRKIGAFYERLLGLEIGENISSVEKSSDISTILKQVCFENNNL
                STETTITKKNGEKYPAEVFVKEISDIHSNSIGIMIIVRDITDQIRLKEMNIELQKKSE
                LEQKRNQELMEARDLALTATKIKSQFLATISHEIRTPLNGIITMGEMLLSTSPLNTEQ
                HDIAETIFGSSELLLSITSDILDFSKIEASKLELEMIEFDFIGCLEGIGKTIGVSITN
                KPIEIAFLMDTDIPHRLIGDPNRLIQIMLNMGTNAVKYTDRGHIVFRISVISREQNRC
                KIKISIDDSGIGIGEDQRAHLFEPFHQIDSSSTRKYGGSGLGLAISSKLAKLMGGEVI
                LERSKPGVGSLFSVTLNFEQIGSNTLKSLLPDKSFSRNKTCIILDNYEYTASIASQRF
                DQIFEESNIILVKESTVKKFFELVQEYKINGDVNKIDLKNLDPELLPFFDSSLICIIV
                FHRFMDNLDIFLKYVKDFIEFYKKKIVVALGINHKNFKNLPKRRDFLIFKKPISSTNL
                IKVSNMARKIIPKSSSSSNLIQTISQIDNQQQQQQQQLQQQQEQQQHQQQQLQQEQQF
                VTISPHSDSSEKKTTPKKDRGKYDFNISPLRIDRFGTESTSSPKIKLYSDTSSDSGES
                DEFEFSENLRELKSDGELVVLKNQPYKEQRLSNIQLINDPIVTPPTPKANSEVVSTKA
                ESTFFQSSSITIKETRSDSVNSVNSVNSVNSVNSLNSENSNNSLDVGIRTRPRILLVD
                DNAVNRKVVKLQLKKLGYDCDTATNGFEGFEMQKKDNYELIFMDLNMPLCDGSMASKL

  Query Match             4.1%;  Score 54.4;  DB 2;  Length 331039;
  Best Local Similarity   47.5%;  Pred. No. 1.6;
  Matches  226;  Conservative    0;  Mismatches  246;  Indels    4;  Gaps    2;

Qy          1 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 60
              || | |  |||||||| ||  |   | |   ||     ||| |  | ||||| ||  |
Db     211098 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 211157

Qy         61 AAGA-TAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATC 119
              |  | || |||   |||||  |   ||| | ||| || ||||| ||| |       ||
Db     211158 ATTATTATTATTATTATTATTGTTATTATTATTATTATTGTTATTATTATTATTATTGGT 211217

Qy        120 TAAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATT 179
              |  |    ||       |  |  |    |||     | | || ||  | ||  | || ||||
Db     211218 TTTACTATCATCAACATTATTGTTATTATTATTATTATTATTGTTATTATTGTTATTATT 211277

Qy        180 TGAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAA 239
                |||   | ||| |    ||    ||  |||      |   |||| | | |   ||||
Db     211278 GTTATTATTGTTATTATTATTACCATTTTGATTATTATGGTTTGAATTATTA---TTAAT 211334

Qy        240 ATATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATC 299
              |    |  ||    | | |  ||  |           |    |  |    |  |||
Db     211335 GTTATCTATTTTAATATGTGGAATATTGTTATTGTAGTTATGGGAATATACAAAATTATT 211394

Qy        300 GTTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGT 359
              |||||    ||   | ||  ||  |    | || ||   ||  || || ||||| || |
Db     211395 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAAT 211454

Qy        360 CGGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGG 419
              |    |  ||   || |   | || |||   |||   |||  || |  ||    ||
Db     211455 TGAATTGAGTGGAGTTGATGCGATAACACAATTGGAATAATTATTTGGAGATTCATCAGT 211514

Qy        420 GGCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTAC 475
              |  || |||||||||| | |  |   |  ||| |||  ||    ||  ||| | |
Db     211515 GAAATTGTTATTATTAAAATTTTTATTAATTTTAAATGGTTGGAGTGGCTCTCTTC 211570


RESULT 15
AR579680/c
LOCUS       AR579680               1141 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 22 from patent US 6784342.
ACCESSION   AR579680
VERSION     AR579680.1  GI:56583130
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1141)
  AUTHORS   Kunst,L. and Clemens,S.
  TITLE     Regulation of embryonic transcription in plants
  JOURNAL   Patent: US 6784342-A 22 31-AUG-2004;
            The University of British Columbia; Vancouver;
            CAX;
FEATURES             Location/Qualifiers
     source          1. .1141
                     /organism="unknown"
                     /mol_type="genomic DNA"
```

ORIGIN

Query Match 4.0%; Score 54; DB 6; Length 1141;
Best Local Similarity 10.9%; Pred. No. 18;
Matches 101; Conservative 336; Mismatches 479; Indels 7; Gaps 2;

```
Qy       327 TGCAGTAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGTATGTTTTTTTTAC 386
Db      1049 WSCATKBMWWTMKWYATKYRTAWYAMWCAWRNNNMWCATNGYAKSCATNNAMWYATTRWA 990

Qy       387 GCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGGTTATTATTAACAATAGTTAC 446
Db       989 AYAAAKWARWAGNNMRMYGAAAGNKWGCMAAMATMGBWWADTAGKMCNNNNNNWTTDVRR 930

Qy       447 AGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGGTTAATTAC 506
Db       929 MAMKAKNNNNNNAYWTACYNRAATNNKMATHWMKWTHGAHSKRRTRHHTRTCRRTKYNNN 870

Qy       507 TGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATATTGGCAGTAAGCTT 566
Db       869 NNNARTVYWYHHAARRWMNAWWTRTNNNNNNNNNNACRNTRTWWABWKHSWCNNNNNNNN 810

Qy       567 ACTTA------TTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTTCAGGGAC 620
Db       809 NNNNNNTWCHYTTANABBCYRANNNNAAARMARTCNNYMHAAVTTTHTDWCYKTWMNTWY 750

Qy       621 TATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAA 680
Db       749 WDMMTTMBTTTTRNMTTSTNMTNNNNNNMWACTNNNNNNMWKAYYAHATNNWGCWWNNTD 690

Qy       681 GCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCTATGCA 740
Db       689 ARRTNNTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNNNNNNNNNNNNNNNSCCTCTR-M 631

Qy       741 ATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTATGACCCCCAT 800
Db       630 MTMRWTMKGDGMTVRKKVKWRDTTCTYVDVWADSWVWWYANWMRCRDVTYTRNNTYCKSY 571

Qy       801 TGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTTTAACATTATG 860
Db       570 AHSYWYWSNNAMWYRRYSARNWSSMARWTTRNNWWMSGBVRMRWAGTMWWRHWNNNNTDT 511

Qy       861 CAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACTAAATATCCAAAATGGGGAGT 920
Db       510 RYYWWWKRWARBTTTVYDSMCNAKSMWRGNNWRAMKMWWAANNDAGAMDHWTYWMGNNTM 451

Qy       921 AATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCGTTCAAAAGTGCAGTAATGCA 980
Db       450 WMRRAWKMMNMAWCRRAYCCNNNNNRACVWHKHKMWRWTWKYMWKAACNNNNBKAMYMRV 391

Qy       981 ATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATACATTTCATTATGGTTATTGAA 1040
Db       390 AWMMYSRDTTNTDWMMWTSDWBWHWYTVDYTMMRAWNNNNNNNNWRBCKTTSWMWWMDHMN 331

Qy      1041 ATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTGCAGGATTTTTTCTTTTTATC 1100
Db       330 THCTYGNNTWGSAYBMAAMSMWAAGASNBVTYNWCWRMTYMGKTMTNNNNNNKAWYYRTK 271

Qy      1101 TTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCAACCTATATAAGAGCGCATAA 1160
Db       270 TVAWCNNRYYYDTAVWTBKRNYKYCYAYBWYYBMYMGKHHWBWWRRABHRSWNMWWVKCR 211

Qy      1161 AACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTCTTGACTATAACTACAATGTT 1220
Db       210 NKYMVSWHYHAMRYBKWABAVGCNNNWKDRMAHHHWCATNNNMMWWWYAYMHHMHKKGKA 151

Qy      1221 GTTTGTTGCATATTTAGAGTACT 1243
Db       150 AWTNNKTABRDDHBAHVKTYWYW 128
```

Search completed: May 11, 2006, 19:14:50
Job time : 4832.84 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 07:09:59 ; Search time 611.03 Seconds
(without alignments)
14626.705 Million cell updates/sec

Title: US-09-423-093B-3_COPY_1019_2359
Perfect score: 1341
Sequence: 1 atgaatcgtattattagaat.....tcaagagttcttatgagtaa 1341

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1341	100.0	12441	2	AAX06750	Aax06750 Salmonell
2	63.8	4.8	15390	14	AEA33353	Aea33353 E. coli O
3	56	4.2	12220	14	AEA88828	Aea88828 Escherich
4	50.6	3.8	19131	4	AAS46717	Aas46717 Tumour su
5	50.4	3.8	8923	4	AAS46340	Aas46340 Tumour su
6	50.2	3.7	17234	6	ABQ67018	Abq67018 Human ang
7	49.8	3.7	21537	6	ABL33999	Abl33999 Human imm
8	49.4	3.7	7255	6	ABL33278	Abl33278 Human imm
9	49.4	3.7	12138	6	ABK40033	Abk40033 Human che
10	49.4	3.7	12138	6	ABL33628	Abl33628 Human imm
11	49.2	3.7	8123	6	ABL34227	Abl34227 Human imm
12	49	3.7	32392	6	ABL56203	Abl56203 AmEPV gen
c 13	48.8	3.6	2000	11	ACL35887	Acl35887 Rice stre
14	48.6	3.6	2000	8	ADA71938	Ada71938 Rice gene
15	48	3.6	1162	14	ADZ71022	Adz71022 Human chr
16	47.8	3.6	2000	11	ACL35363	Acl35363 Rice stre
c 17	47.6	3.5	2000	11	ACL37108	Acl37108 Rice stre
18	47.6	3.5	6577	6	ABL33356	Abl33356 Human imm
19	47.6	3.5	6577	6	ABL70561	Abl70561 Chemicall
20	47.6	3.5	6577	6	AAS61221	Aas61221 Human gen
21	47.6	3.5	14924	6	ABL32224	Abl32224 Human imm
22	47.6	3.5	14924	6	ABL54321	Abl54321 Chemicall
23	47.2	3.5	2501	14	ADV98124	Adv98124 Bisulfite
24	46.6	3.5	6191	6	ABL33217	Abl33217 Human imm
25	46.6	3.5	6191	6	ABK31307	Abk31307 Signal tr
26	46.6	3.5	6191	6	ABL70282	Abl70282 Chemicall
27	46.6	3.5	6191	6	ABN80161	Abn80161 Human che
28	46.4	3.5	6392	6	ABL32684	Abl32684 Human imm
29	46.4	3.5	6392	6	ABL34506	Abl34506 Human met
30	46.4	3.5	6392	7	ADS99767	Ads99767 Bisulphit
31	46	3.4	1452	14	ADZ71095	Adz71095 Human chr
32	46	3.4	15373	6	ABL32467	Abl32467 Human imm
33	45.8	3.4	7625	6	ABL33431	Abl33431 Human imm
34	45.8	3.4	18281	6	ABL32962	Abl32962 Human imm
35	45.8	3.4	71997	13	ABD33343	Abd33343 Human can
36	45.6	3.4	3720	6	ABK33954	Abk33954 Human DNA
37	45.6	3.4	3720	8	ABZ10006	Abz10006 Haematopo
38	45.6	3.4	3720	8	ADA20409	Ada20409 Prostate
39	45.6	3.4	3720	8	ADA84216	Ada84216 Human ren
40	45.6	3.4	3720	10	ADB54084	Adb54084 Pretreate
41	45.6	3.4	3720	10	ADE84070	Ade84070 Human lym
42	45.6	3.4	3720	13	ADS89238	Ads89238 Oligonucl
43	45.6	3.4	8030	6	ABL33333	Abl33333 Human imm
44	45.6	3.4	17869	6	ABK39920	Abk39920 Human che
45	45.6	3.4	17869	6	ABL32104	Abl32104 Human imm

ALIGNMENTS

RESULT 1
AAX06750
ID AAX06750 standard; DNA; 12441 BP.
XX
AC AAX06750;
XX
DT 26-APR-1999 (first entry)
XX
DE Salmonella enterica serovar C2 O antigen gene cluster.
XX
KW O antigen; diagnosis; wbaR gene; wzx gene; wbaL gene; wbaQ gene;
KW wxy gene; wbaW gene; wbaZ gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Salmonella enterica.
XX
FH Key Location/Qualifiers
FT CDS 138. .1019
FT /*tag= a
FT /note= "abe gene; encodes AAW88324"
FT CDS 1019. .2359
FT /*tag= b
FT /product= "flippase"
FT /note= "wzx gene, this region is specifically claimed in
FT Claim 9; encodes AAW88325"
FT CDS 2362. .3314
FT /*tag= c
FT /product= "abequosyl_transferase"
FT /note= "wbaR gene, this region is specifically claimed in
FT Claim 9; encodes AAW88326"
FT CDS 3361. .3975
FT /*tag= d
FT /product= "acetyl_transferase"
FT /note= "wbaL gene, this region is specifically claimed in
FT Claim 9; encodes AAW88327"
FT CDS 3977. .5020
FT /*tag= e
FT /product= "rhamnosyl_transferase"
FT /note= "wbaQ gene, this region is specifically claimed in
FT Claim 9; encodes AAW88328"
FT CDS 5114. .6313
FT /*tag= f

```
FT                   /product= "O_antigen_polymerase"
FT                   /note= "wzy gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88329"
FT   CDS             6313. .7323
FT                   /*tag= g
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88330"
FT   CDS             7310. .8467
FT                   /*tag= h
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88331"
FT   CDS             8482. .9903
FT                   /*tag= i
FT                   /note= "manC gene; encodes AAW88332"
FT   CDS             9911. .11347
FT                   /*tag= j
FT                   /note= "manB gene; encodes AAW88333"
FT   CDS             11426. .12441
FT                   /*tag= k
FT                   /note= "wbaP gene; encodes AAW88334"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88324, AAW88325, AAW88326, AAW88327, AAW88328, AAW88329,
DR   AAW88330, AAW88331, AAW88332, AAW88333, AAW88334.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 9; Fig 9; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Salmonella enterica serotype C2 (strain M67) O
CC   antigen. 11 Open reading frames within the gene cluster were identified
CC   on the basis of homology to known sequences. The genes encode proteins
CC   (see AAW88323-34) involved the synthesis of sugars present in the
CC   polysaccharide antigen, and in the transport or processing of
CC   polysaccharide or oligosaccharide units. The use of nucleic acid
CC   molecules derived from assembly and transport genes, particularly wbd
CC   (transferase), wzx (flippase) and wzy (polymerase) genes, within O
CC   antigen gene clusters improves the specificity of methods for the
CC   detection and identification of O antigens, e.g. in testing food- or
CC   faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of S. enterica
XX
SQ   Sequence 12441 BP; 3887 A; 1876 C; 2518 G; 4160 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1341;  DB 2;  Length 12441;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-291;
  Matches 1341;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 1078

Qy         61 AAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 AAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCT 1138

Qy        121 AAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 AAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTT 1198

Qy        181 GAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 GAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAA 1258

Qy        241 TATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 TATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCG 1318

Qy        301 TTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 TTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTC 1378

Qy        361 GGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 GGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGG 1438

Qy        421 GCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 GCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCT 1498

Qy        481 GTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 GTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTG 1558

Qy        541 TTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 TTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCT 1618

Qy        601 GCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1619 GCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAA 1678

Qy        661 ATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 ATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTT 1738

Qy        721 AATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 AATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATT 1798

Qy        781 TATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 TATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTA 1858

Qy        841 GGGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 GGGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACT 1918

Qy        901 AAATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1919 AAATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCG 1978

Qy        961 TTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1979 TTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATAC 2038

Qy       1021 ATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2039 ATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTG 2098

Qy       1081 CAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2099 CAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCA 2158

Qy       1141 ACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2159 ACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTC 2218
```

```
Qy       1201 TTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACATGTTA 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2219 TTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACATGTTA 2278

Qy       1261 ATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTTAAAAGA 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2279 ATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTTAAAAGA 2338

Qy       1321 TTCAAGAGTTCTTATGAGTAA 1341
             |||||||||||||||||||||
Db       2339 TTCAAGAGTTCTTATGAGTAA 2359
```

RESULT 2
AEA33353
ID AEA33353 standard; DNA; 15390 BP.
XX
AC AEA33353;
XX
DT 28-JUL-2005 (first entry)
XX
DE E. coli O156 O-antigen sequence, seq id 1.
XX
KW Microorganism detection; O-antigen; gene; ds.
XX
OS Escherichia coli; O156.
XX
PN CN1546512-A.
XX
PD 17-NOV-2004.
XX
PF 03-DEC-2003; 2003CN-01107162.
XX
PR 03-DEC-2003; 2003CN-01107162.
XX
PA (UYNA-) UNIV NANKAI.
XX
PI Feng L, Kong Q, Guo H;
XX
DR WPI; 2005-164294/18.
XX
PT Polynucleotide specific for Escherichia coli 0156 O-antigen.
XX
PS Claim 1; SEQ ID NO 1; 31pp; Chinese.
XX
CC The invention provides a polynucleotide specific for the Escherichia coli
CC 0156 O-antigen, which is the total nucleotide sequence of the gene
CC cluster for controlling the synthesis of O-antigens in E. coli 0156. It
CC includes the glycosyl transferase genes and oligosaccharide unit
CC treatment genes in the O-antigen gene cluster originated from E. coli
CC 0156. The invention includes detecting and identifying E. coli 0156 in
CC the human body and in the environment (e.g. by PCR). The current sequence
CC represents the E. coli O156 O-antigen sequence.
XX
SQ Sequence 15390 BP; 4762 A; 2581 C; 3133 G; 4914 T; 0 U; 0 Other;

Query Match 4.8%; Score 63.8; DB 14; Length 15390;
Best Local Similarity 47.5%; Pred. No. 0.00034;
Matches 373; Conservative 0; Mismatches 382; Indels 30; Gaps 5;

```
Qy         74 TATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCTAAAGACGCACAGG 133
              ||||||  ||      |  ||||  |  |  | ||| |    |||  |  ||| | |   |
Db       1274 TATTAAGTGGACCCATAATAATTCTTGTTGTTTCAAAGACATTAAGTTCAGAGGAAATTG 1333

Qy        134 GCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTTGAATTGGGGCTAT 193
              | | |||||||||| |||       ||||  |  | |||    ||  |||||| ||    |
Db       1334 GATTTTATTATACTTTTTTCAGCTTAGTAACTATGCAGCAGTTAGCTGAATTAGG----T 1389

Qy        194 CAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTATTCTG 253
                | || || |||| |  | |    |||        |||||  |||   ||  | || |
Db       1390 TTAGGGCAAACATTAAGATTG-----CATATTGCTCATGCGTTTAAAATGGACGAGTCGG 1444

Qy        254 AACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCGGTTGG 313
              | |   | |||       |||   ||| |||||||           |  | |    |
Db       1445 GATGGTGTTTTAACTCAAAAAAAGAAATTAAGCAA-----------TACTATTCTCTAA 1492

Qy        314 CAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGT 373
              |||||||||      ||  | || ||    |  ||  | ||  | |||||     ||
Db       1493 GTGTAAAATGGTTCTTGGTCGTTGCGTTATGTACTATCATTATTATTGGTCCTGCGGGTA 1552

Qy        374 ATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGGTTATTAT 433
              |  || || |        |  ||||    | | |   |||| |||| |||||| ||| ||
Db       1553 TTTATTATTATCATTCTTACCAAGG---GGATATTATTTGGGAAGGCCCATGGGTATCAT 1609

Qy        434 TAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGA 493
              ||  | | |||||||||  | |  ||| |  |      || |||           ||||
Db       1610 TAGTATTGGTTACAGCACTAACAATTATAATAACGCCTATACAAATCACAATTGAATCTC 1669

Qy        494 GTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATAT 553
                       | |   ||  ||||      |  |||   ||  |  |  | ||||||
Db       1670 TACAACATCAAATTTCAGTATTTAAAGCACAGACATATGCATCAATTATCAATAGTATAT 1729

Qy        554 TGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTT 613
                 |      ||| |  ||  ||| ||||| |||||||| |   ||  ||       ||
Db       1730 CATTATGTCTGTTATTGTTTTCTGGATTTGGTCTATATGCAATTGCTTTAAGCCTATTTA 1789

Qy        614 CAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAAT 673
                 | |   | | | |||  |||||||  ||     |      ||||  ||   |  |||  |
Db       1790 TCAGTAACTTAACAGTCTTAATATTTTTATGGGGG------AAAATAAAATGGCTGAGAA 1843

Qy        674 CTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTC 733
                |      |       |   |  |  |  |  | ||       ||||  || || |  |
Db       1844 ATAACCTGATGCCTGTTCTTGTTGATACAAATTTTTTTGTTGTTTTTAAAGAGATCTGGC 1903

Qy        734 CTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTATGA 793
              |  ||   |  ||  | ||| |    |   | |  || |   || |||    ||   | |
Db       1904 CGCTGTTTTCTCGTGTAGCTGTTGTATCTGTATTTGGTTTCCTTTTTTGGAATGGATTTA 1963

Qy        794 CCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTTTAA 853
                |  |||||||| |||   |  | | |   |||||||| ||  ||    | ||  |||
Db       1964 ATCTAATTGCATTTAAAGTATATGAGCCCTCATATGCAGGACAAGTAATTTTATCCATAA 2023

Qy        854 CATTA 858
                | |||
Db       2024 CCTTA 2028
```

RESULT 3
AEA88828
ID AEA88828 standard; DNA; 12220 BP.
XX
AC AEA88828;
XX
DT 11-AUG-2005 (first entry)
XX
DE Escherichia coli O18 O-antigen biosynthetic gene cluster.
XX
KW Microorganism detection; O-antigen biosynthetic gene cluster;
KW lipopolysaccharide antigen; gene; ds.
XX
OS Escherichia coli; O18.
XX

FH	Key	Location/Qualifiers
FT	CDS	1415. .2434
FT		/*tag= a
FT		/product= "Orf1 protein"
FT	CDS	2800. .3171
FT		/*tag= b
FT		/product= "Orf2 protein"
FT	CDS	3407. .4567
FT		/*tag= c
FT		/product= "Wsx protein (orf3)"
FT	CDS	4551. .5723
FT		/*tag= d
FT		/product= "Orf4 protein"
FT	CDS	5720. .6526

```
FT                   /*tag=  e
FT                   /product= "Orf5 protein"
FT   CDS             6516. .7745
FT                   /*tag=  f
FT                   /product= "Wsy protein (orf6)"
FT   CDS             8021. .8755
FT                   /*tag=  g
FT                   /partial
FT                   /product= "Orf7 protein"
FT                   /note= "No start codon given"
FT   CDS             8800. .9846
FT                   /*tag=  h
FT                   /product= "Orf8 protein"
FT   CDS             9843. .10631
FT                   /*tag=  i
FT                   /product= "Orf9 protein"
XX
PN   CN1554759-A.
XX
PD   15-DEC-2004.
XX
PF   22-DEC-2003; 2003CN-01117863.
XX
PR   22-DEC-2003; 2003CN-01117863.
XX
PA   (UYNA-) UNIV NANKAI.
XX
PI   Wang L,  Kong Q,  Feng L;
XX
DR   WPI; 2005-215827/23.
XX
PT   O-antigen specific nucleotide of E. coli 081 type.
XX
PS   Claim 1; SEQ ID NO 1; 29pp; Chinese.
XX
CC   The invention relates to the O-antigen biosynthetic gene cluster of
CC   Escherichia coli O81 (AEA88828) and to variants thereof. The O-antigen of
CC   E. coli O81 is a lipopolysaccharide antigen. The invention also relates
CC   to PCR primers (AEA88829-AEA88830) for the amplification of the gnd gene
CC   of the O-antigen gene cluster, oligonucleotides derived from the
CC   glycosyltransferase and oligose unit processing genes of the O-antigen
CC   gene cluster, and a method for detecting E. coli O81 in the environment
CC   and in the human body using oligonucleotides of the invention. The
CC   present sequence represents the specifically claimed sequence of the E.
CC   coli O81 O-antigen biosynthetic gene cluster.
XX
SQ   Sequence 12220 BP; 3952 A; 1820 C; 2307 G; 4141 T; 0 U; 0 Other;

  Query Match             4.2%;  Score 56;  DB 14;  Length 12220;
  Best Local Similarity   44.0%;  Pred. No. 0.018;
  Matches  337;  Conservative    0;  Mismatches  420;  Indels    9;  Gaps    2;

Qy       259 GATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCGGTTGGCAATA 318
Db      3494 GCTATTCATGATATCAGTGAAAAAAATATAAAAATATCGAACTATTTTAACTTTTCAATC 3553

Qy       319 AAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGTATGTT 378
Db      3554 TATTGGTATTTAATATTATCAATACTTTTTTTTATATTGGGAGGGGGGGTGTCATATGTT 3613

Qy       379 TTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGGTTATTATTAACA 438
Db      3614 ATGTTTAATGATTATAACGGAATA---GTTAACTGGAAATTACCATGGTTTTTTTTGCTT 3670

Qy       439 ATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGG 498
Db      3671 TTTGTTTCATGTACTAATTTATTAGTATTGCCCTTAAATGCAATGCTTGATGGAACACAA 3730

Qy       499 TTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATATTGGCA 558
Db      3731 AGGCAACAATTATTGATGAAAGCTAATATTGTTAGTCAGTTAATGATAGCGATATCATTA 3790

Qy       559 GTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTTCAGGG 618
Db      3791 TGGCTGACAATTTATTTAGGCTTGGGGCTATATTCTCTTGGTCTATCACAATTGGTAGGA 3850

Qy       619 ACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTA 678
Db      3851 ATATTATTTTTTTTACTGTCAGTTTTTATATTTAATAAATGTAGATTCCAGTTCATAGAA 3910

Qy       679 AAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCTATG 738
Db      3911 ATACATAAAAAAAACTTTTC------ATTTAAAACAGTATTTAATGAGTTGTGGCCGTTA 3964

Qy       739 CAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTATGACCCCC 798
Db      3965 TTGAAAAAAACTAGTATCGTATGGTTTCTTGGTTATTTCTACTGGAACGGTTTTAATATA 4024

Qy       799 ATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTTTAACATTA 858
Db      4025 ATATCATTTAAATATCTTGGTGCAGAGATCGCAGGCTTCGTTGGTATGTCGATGTCAATT 4084

Qy       859 TGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACTAAATATCCAAAATGGGGA 918
Db      4085 ATGCGAGCGGGACAGAATATAGCAATATCTATTTTGAATTCTCAGATGACATTGTATGCA 4144

Qy       919 GTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCGTTCAAAAGTGCAGTAATG 978
Db      4145 AGCAATATTGCGATGGGAAAGATTGATGAAGCAAGAAGAGTATTTAAGAAATATTTTATC 4204

Qy       979 CAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATACATTT 1024
Db      4205 ATTTCATTTTTGTTATTAATTTCTGGTTACTCAATATTTTATATTT 4250

RESULT 4
AAS46717
ID   AAS46717 standard; DNA; 19131 BP.
XX
AC   AAS46717;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #440.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 440; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
```

CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 19131 BP; 5767 A; 407 C; 3981 G; 8976 T; 0 U; 0 Other;

Query Match 3.8%; Score 50.6; DB 4; Length 19131;
Best Local Similarity 45.9%; Pred. No. 0.33;
Matches 173; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

Qy 136 TATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTTGAATTGGGGCTATCA 195
Db 72 TATTGTAAAAAGAAAAAGAAAATGATGTTAATAAAAAGTATATTGTATTGTTATAAGTTA 131

Qy 196 ACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTATTCTGAA 255
Db 132 TATAAAAATAAAAATATAGTTATAAATAATAATTTATTGTGAAGATTATATTTTTTATAA 191

Qy 256 CGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCGGTTGGCA 315
Db 192 AGTTTATGTATGGTTTAAGTTAAAATTGTAAATAGTTATAGTTTTTTTGATAGATTATTA 251

Qy 316 ATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGTAT 375
Db 252 AATATGATAAAATTATGTAAAAAGTAGGTAATGTTATTTATGATTTTTAATATTAGTGAA 311

Qy 376 GTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGGTTATTATTA 435
Db 312 TTTTTATTTAAAAAGAAAGAAGGAAAAGAAAAATAAGTGATACGTATTTTTGGATTGTTT 371

Qy 436 ACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGT 495
Db 372 AAGATATTTAAAATATTTTAAATAAATTTTATTTTTTTATAGTATTAGAATATAGAAAGG 431

Qy 496 GGGTTAATTACTGATGT 512
Db 432 TAATTTTTTATTGTTTT 448

RESULT 5
AAS46340
ID AAS46340 standard; DNA; 8923 BP.
XX
AC AAS46340;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #62.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 62; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 8923 BP; 2227 A; 270 C; 2397 G; 4029 T; 0 U; 0 Other;

Query Match 3.8%; Score 50.4; DB 4; Length 8923;
Best Local Similarity 50.0%; Pred. No. 0.31;
Matches 126; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

Qy 621 TATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAA 680
Db 1228 TATTGGATTATTTTTTTTTTTTGTATATATTTTTAATTTAAAATTTTGAATTATTAAGA 1287

Qy 681 GCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCTATGCA 740
Db 1288 GGAATAAATTAATTGTTTAAAGGTTAATTTATAATGAAAAAAATAATTTGTTTATGATAA 1347

Qy 741 ATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTATGACCCCCAT 800
Db 1348 TTTTAAAGGTATTTTTAATAGTATTATAAGTTATATTTTTTTTGATATTAAAATTAAATG 1407

Qy 801 TGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTTTAACATTATG 860
Db 1408 TAGATTATTTTTTTTATGTGGTATTTATAAAATAAATATAGTAATGAATTATTTAATATT 1467

```
Qy        861 CAATATGGTAAT 872
              ||||| ||||
Db       1468 AAATATTTTAAT 1479


RESULT 6
ABQ67018
ID   ABQ67018 standard; DNA; 17234 BP.
XX
AC   ABQ67018;
XX
DT   28-AUG-2002  (first entry)
XX
DE   Human angiogenesis associated polynucleotide SEQ ID NO 48.
XX
KW   Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW   inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW   macular degeneration; inflammatory bowel disease; Crohn's disease;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiarteriosclerotic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200246454-A2.
XX
PD   13-JUN-2002.
XX
PF   06-DEC-2001; 2001WO-EP014320.
XX
PR   06-DEC-2000; 2000DE-01061338.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Schacht O;
XX
DR   WPI; 2002-500450/53.
XX
PT   New nucleic acid fragments from chemically treated angiogenesis-
PT   associated genes, useful for determining methylation status, e.g. in
PT   diagnosis or treatment of cancer.
XX
PS   Claim 1; SEQ ID NO 48; 41pp + Sequence Listing; German.
XX
CC   The invention relates to a nucleic acid (I) comprising a segment of 18
CC   bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC   having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC   related oligomers, are used to evaluate the methylation status and/or
CC   single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC   diagnosis and treatment of eye diseases, proliferative retinopathy,
CC   neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC   diabetic retinopathy, macular degeneration caused by neovascularisation,
CC   psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC   Crohn's disease. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 17234 BP; 4664 A; 169 C; 3856 G; 8545 T; 0 U; 0 Other;

  Query Match             3.7%;  Score 50.2;  DB 6;  Length 17234;
  Best Local Similarity   43.6%;  Pred. No. 0.4;
  Matches  328;  Conservative    0;  Mismatches  418;  Indels    7;  Gaps    2;

Qy         41 TTCGTTATGTTATTTTTGGTAAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGT 100
              || ||    |  ||| ||   | | |  ||| |||  ||  |||  |  |  |    || |
Db       2296 TTTGTACGGAAATTATTTAGATGTTTTTATGTGTTGTAGAGGGGATGTTAGTTGAGATTT 2355

Qy        101 TAATATCACACCATTTATCTAAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAG 160
               | |  |      || |   ||     |  |   |||| ||  |   |||| |
Db       2356 AGAAAATATTTTTGTTTTGGTAATTGTTAAGGTTTTATTTTTTAAATGTTAAGTTTTAGA 2415

Qy        161 TGGCACTACAGATAATATTTGAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCC 220
              |   | ||   ||  |||  ||  |   |     |    |  ||  |     ||
Db       2416 TTTTAATATTTTTATGATTATAAGGGATTTTGGAGTAGTATTTTTTTTGAGGATTTAAAA 2475

Qy        221 ATGAAATGTCAGCGTTAAAATATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAA 280
              | ||||||  |   |||| | | | ||  |        || || |||    |   | |||
Db       2476 TTAAAATGTTATTTTTAATAGAGGTTTTGTGGTGGAGATATTTTATTTTTTGTATTAATT 2535

Qy        281 ATAAGCAACGTTACCTATCGTTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTT 340
              | | |     |||  | |   |  ||            |    |  |||  | | | | |
Db       2536 AAATGGTTTTTTATTTTTAAGTGATTTTAGAAAGTTTTTTTTTTTATTGATTTAAATGTG 2595

Qy        341 TGCTAATAATATTAATAGTCGGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCT 400
              |  |  | | |||  | |          ||    | | |||| |       ||| |  |  |
Db       2596 TTTTTTTTAAATTTTTTGATTTGTTTATTTTATTTTTTTTTGTAGTTATAAAGAATAAATT 2655

Qy        401 TAGGTGTACCTTGGCAAGGGGCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTT 460
              |  | |    |    |    | ||||||  |  ||||||      |||  | |||
Db       2656 TGTTTTTTTTATTTATAAATAATTTTTTATTAGGAGTAATAGTGGTTATGTTTTTTTTTT 2715

Qy        461 TTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAA 520
              || || |||    | ||| |    ||   ||  |    |  ||  |  | |  | |
Db       2716 TTTTTTTTTTAAAATTTTTT-----TGTTAGGTTTATTATTTTTTATTTTTTTTAAGATT 2770

Qy        521 TGAGAATGTATCAGTCGCTGTTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCT 580
              |   |||  || |   | || | |  |||| | || | || |   |   ||   | | |
Db       2771 TTTTAATTGATTAAATGGTGGTGGGAGGTAATTCGGTATTATGGAGAAAAATATTTAGTT 2830

Qy        581 TTGGACTATATGCTACGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTT 640
              |   | || ||| || |          |||  ||     | |  || |   | ||||
Db       2831 TGAAATTAAATGTTAGGGGGTTTTGTTTTATTATAGTAGTATTAAGTTTTTTGGTTTTTT 2890

Qy        641 CATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTG 700
              | ||     ||||   |    ||   || |||   || | |    |||    |   |
Db       2891 TAGATTTTGGTTTTTTTATTTATTAATTGAAATGGGTAGATTAGGTGAATTTTGAAGTGG 2950

Qy        701 AAGGT--GGTATTTCATGGGTTAATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAG 758
              || |  ||| ||| |   |||  | | |||||     |  | |   |  ||| | ||
Db       2951 TAGTTTCGGTTTTTAAAATGTTTTTAAGATATTTTAAGTATATTTTGGTTTTGATATATA 3010

Qy        759 TTGGATGTCAGGGTATTTTATTTATTTTGTTAT 791
              | |  | |   || ||| ||| | |||  || |
Db       3011 TGGTTTTTTTTGGGATTATATGTTTTTATTTGT 3043


RESULT 7
ABL33999
ID   ABL33999 standard; DNA; 21537 BP.
XX
AC   ABL33999;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1972.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
```

XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1972; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 21537 BP; 6538 A; 333 C; 4426 G; 10240 T; 0 U; 0 Other;

Query Match 3.7%; Score 49.8; DB 6; Length 21537;
Best Local Similarity 46.6%; Pred. No. 0.51;
Matches 159; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

```
Qy        345 AATAATATTAATAGTCGGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGG 404
              ||||  || ||| ||    |   |      || |    |||| |      ||        ||
Db        146 AATAGGATAAATTGTTTTTTTTAATAATAATATAAAATTTATGATTTTAGTTAATATATA 205

Qy        405 TGTACCTTGGCAAGGGGCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCT 464
              | |   || | |     |  |  || || ||   | ||   |    | |||   ||| |
Db        206 TATTTTTTAGAAGATAGATTTTTTTTTTTTGTTATTAAATTGTGTTTTGGGTATGTAAAG 265

Qy        465 TGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAG 524
              |       ||| ||| |||       | ||  || |||  ||  |||| | | | || |||
Db        266 TTAAAGGGTAGTTTTTGTGATATTATGGTTTGAGTTTTTTTATGATTTAATTTAAGTGAT 325

Qy        525 AATGTATCAGTCGCTGTTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGG 584
              | |        |   |||||      ||||||  |||| | ||     ||| | |   |
Db        326 ATTTATATTATTATTGTTATTGTTAATATTGTTAGTAGGTTTTTAGTTTATTTGGAATTT 385

Qy        585 ACTATATGCTACGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATA 644
              |  | |    |          || | | |   | | ||   |        |||||| | |
Db        386 ATAAAAGTAAAAAATGTATTAAGTTTTGTAGAGTATTAGGGTTAAAAATATATTTTTAAA 445

Qy        645 TAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAAGCATA 685
              | | | |||||| ||||   |  || ||  |||||   |||
Db        446 TGATTGTTTTAATAAAAGCGTTATGAAAAATTTAATTTATA 486
```

RESULT 8
ABL33278
ID ABL33278 standard; DNA; 7255 BP.
XX
AC ABL33278;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1251.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1251; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 7255 BP; 2118 A; 40 C; 1359 G; 3738 T; 0 U; 0 Other;

Query Match 3.7%; Score 49.4; DB 6; Length 7255;
Best Local Similarity 48.1%; Pred. No. 0.49;
Matches 140; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

```
Qy        501 AATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATATTGGCAGT 560
              || |  | || |  |||  |  | || | ||  || | | |    | |||| ||   | |
Db       3382 AAGTTGTTATATTTATATTAATATAAAGAATGGGTTGTTATATTTTTGTATGTTTATATT 3441

Qy        561 AAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTTCAGGGAC 620
                ||  |||| | | | |||  | | |||  |   | || |||   | ||      |
Db       3442 TTATTTTTTTATGATTCGGTTTAAATTTTATTTTTTTTATGAAATTTTTTTTGATTTTAT 3501

Qy        621 TATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAA 680
              | |  |  | |   | ||  | |    || || | |  |||||  |  ||| ||||
Db       3502 TTTTTTTATATTATTGTTGTTTTTTTTAAATATTTATAGTATTTTATTTAAATTTTTATT 3561

Qy        681 GCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCTATGCA 740
                | |   | |||||  |    |  |  |     |    ||| | ||  |  || | |
Db       3562 AATTTATTTTTAATATTTTAGTATTTATTATAGTATGTGGTATGTAGTAGATGATAGGAA 3621

Qy        741 ATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTAT 791
              |||   | |   |  | || |||    |   ||||    ||| ||||||
Db       3622 ATGTTTATTGAATGAATGTAGGAATATATATTATTAATATTAATTTGTTGT 3672
```

RESULT 9
ABK40033
ID ABK40033 standard; DNA; 12138 BP.
XX
AC ABK40033;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #58 strand 1.
XX
KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO200202806-A2.

XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007470.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 115; 24pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 12138 BP; 3002 A; 183 C; 2766 G; 6187 T; 0 U; 0 Other;

Query Match 3.7%; Score 49.4; DB 6; Length 12138;
Best Local Similarity 48.7%; Pred. No. 0.55;
Matches 163; Conservative 0; Mismatches 171; Indels 1; Gaps 1;

```
Qy       957 ATCGTTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGT 1016
Db        99 ATTGGTAGAGAGAGTATTTTTATATTTATTTTTTTGTTTGTAGAGTAGGAGGAGTGGTAT 158

Qy      1017 ATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGG 1076
Db       159 TTGGAATTGATTTATGGTTTTAGTAATTATTAGTTTTTTAAAGGGTTTTAAGTTTTTAGA 218

Qy      1077 ATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTT 1136
Db       219 TTAGTTTGTGTGTGTGTATGTTTAGTGTGCGGTGGTATGTTTTTATAATTTAGTTTTGTA 278

Qy      1137 TGCAACCTATATAAGAGCGCAT-AAAACTGAAAAAATGACATTGGCATCATGTATAATGG 1195
Db       279 AAGAAGTTTTTTGCGATCGGTTCGTATATAGGTTTTTTAGTTTGGTAGGTTTTATTATTT 338

Qy      1196 CTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACA 1255
Db       339 TTATTTTTATAATAATTTTTATATTTTTAAGTGTTTTGTTAGTATATTTGATTTTTTAAA 398

Qy      1256 TGTTAATGTATGCAGCACTAACGTGGTTATATTTT 1290
Db       399 TAGTTTTGTAAGATAAATTGAATTGGTATTATTTT 433
```

RESULT 10
ABL33628
ID ABL33628 standard; DNA; 12138 BP.
XX
AC ABL33628;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1601.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1601; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 12138 BP; 3002 A; 183 C; 2766 G; 6187 T; 0 U; 0 Other;

Query Match 3.7%; Score 49.4; DB 6; Length 12138;
Best Local Similarity 48.7%; Pred. No. 0.55;
Matches 163; Conservative 0; Mismatches 171; Indels 1; Gaps 1;

```
Qy       957 ATCGTTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGT 1016
Db        99 ATTGGTAGAGAGAGTATTTTTATATTTATTTTTTTGTTTGTAGAGTAGGAGGAGTGGTAT 158

Qy      1017 ATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGG 1076
Db       159 TTGGAATTGATTTATGGTTTTAGTAATTATTAGTTTTTTAAAGGGTTTTAAGTTTTTAGA 218

Qy      1077 ATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTT 1136
Db       219 TTAGTTTGTGTGTGTGTATGTTTAGTGTGCGGTGGTATGTTTTTATAATTTAGTTTTGTA 278

Qy      1137 TGCAACCTATATAAGAGCGCAT-AAAACTGAAAAAATGACATTGGCATCATGTATAATGG 1195
```

```
                ||  | | |  || ||  |   |  |       | |  |||| |   | ||| ||
Db        279 AAGAAGTTTTTTGCGATCGGTTCGTATATAGGTTTTTTAGTTTGGTAGGTTTTATTATTT 338

Qy       1196 CTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACA 1255
               | | || |  |||| |   || || ||   ||  |  ||||  || | ||  || || |
Db        339 TTATTTTTATAATAATTTTTATATTTTTAAGTGTTTTGTTAGTATATTTGATTTTTTAAA 398

Qy       1256 TGTTAATGTATGCAGCACTAACGTGGTTATATTTT 1290
              |  |  |||| |    | | |  ||||  ||||||
Db        399 TAGTTTTGTAAGATAAATTGAATTGGTATTATTTT 433


RESULT 11
ABL34227
ID   ABL34227 standard; DNA; 8123 BP.
XX
AC   ABL34227;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2200.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 2200; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 8123 BP; 2184 A; 235 C; 1860 G; 3844 T; 0 U; 0 Other;

  Query Match             3.7%;  Score 49.2;  DB 6;  Length 8123;
  Best Local Similarity   46.9%;  Pred. No. 0.56;
  Matches  153;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy        988 TTTTTTGTCTTGACAGGATTAACTGGTGTATACATTTCATTATGGTTATTGAAATTATCT 1047
              ||||||  | |||      || |        ||  ||| ||| |  || ||| || ||| |
Db        144 TTTTTTTTTTTGTTGTATTTTAGGATAACGTAATTTTAATTTTTTTTTTTGTAAATATAT 203

Qy       1048 GGTTCAAACATTGGCGAGCGGTTTTTGGGATTGCAGGATTTTTTCTTTTTATCTTTAGCA 1107
                ||     | |   ||  | |||          |  |||||| | ||||  |||
Db        204 TATTTGTTTAGTAAAGACGGTTTTAATTTTAATTCGTTTTTTTTTTTATTTAAATTTTTTT 263

Qy       1108 ATTATTGGTAATCACATTGTAGCTTGCTTTGCAACCTATATAAGAGCGCATAAAACTGAA 1167
                | ||  ||||   || || | ||  ||||     || ||        | |  | |  |
Db        264 TATTTTTTTAATTGTATAGTTGTTTTGTTTGTTTTTTAAATGGAGTATTAAATTAGTTTA 323

Qy       1168 AAAATGACATTGGCATCATGTATAATGGCTCTCTTGACTATAACTACAATGTTGTTTGTT 1227
              ||| | |   |    |       || | | | | | |  | | || |  | ||| ||| ||
Db        324 AAATTAAATGTAATTTAGGTAGTATTTGTTTTTTAGTTTTTTACGATTAAGTTTTTTTTT 383

Qy       1228 GCATATTTAGAGTACTCGAGGTTCTACATGTTAATGTATGCAGCACTAACGTGGTTATAT 1287
              | | || | |      |||| |  | || ||||| |    |  | |    ||| | |
Db        384 GAAAATATCGGTGTGGTGAGGGTTAAGATTTTAATTTTGTTATTAGTTGTTTGGGTTTGG 443

Qy       1288 TTTGTTCCTCAAACTTATATAATCTT 1313
                |  | |   | || |   || ||
Db        444 GTAAGTTATTTTAGTTTTTAGATGTT 469


RESULT 12
ABL56203
ID   ABL56203 standard; DNA; 32392 BP.
XX
AC   ABL56203;
XX
DT   01-JUL-2002  (first entry)
XX
DE   AmEPV genome fragment#5.
XX
KW   AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
KW   genetic deficiency disorder; ds.
XX
OS   Amsacta moorei entomopoxvirus.
XX
PN   WO200212526-A2.
XX
PD   14-FEB-2002.
XX
PF   10-AUG-2001; 2001WO-US025287.
XX
PR   10-AUG-2000; 2000US-0224479P.
PR   14-SEP-2000; 2000US-00662254.
XX
PA   (UYFL ) UNIV FLORIDA.
XX
PI   Moyer RW,  Li Y,  Bawden AL;
XX
DR   WPI; 2002-227161/28.
XX
PT   Novel recombinant entomopox virus vector useful for delivering
PT   polynucleotide encoding protein to vertebrate cell, comprises
PT   polynucleotide encoding protein operably linked with heterologous
PT   promoter sequence.
XX
PS   Disclosure; Page 226-242; 326pp; English.
XX
CC   The invention relates to a recombinant entomopox virus (EPV) vector,
CC   comprising a polynucleotide encoding a protein operably linked with a
CC   heterologous promoter sequence. The invention also concerns methods for
CC   providing gene therapy for genetic deficiency disorders. Vectors of the
CC   invention are useful for delivering a polynucleotide encoding a protein
CC   to a vertebrate cell preferably a mammalian cell, such as a human cell.
CC   The vector is introduced into the vertebrate cell by infection in a viral
CC   particle, or by transfection, transduction, or injection either in vitro
CC   or in vivo. The vector is useful for the delivery and expression of
CC   biologically useful proteins in gene therapy protocols, and for
CC   delivering large DNA segments for engineering of vertebrate cells.
CC   Polynucleotides of the invention have applications in techniques such as
CC   their use as insertion sites for foreign genes of interest, hybridisation
CC   probes, for chromosome and gene mapping, in PCR technologies, and in the
CC   production of sense or antisense nucleic acids. Vectors of the invention
```

```
CC   provide for stable integration and expression of heterologous DNA in host
CC   cells, and are adapted for accepting large heterologous polynucleotide
CC   inserts which can be delivered in an infected or transformed cell and
CC   expressed in a stable fraction. The current sequence represents a
CC   fragment of the genome of the genus B entomopoxvirus from amsacta moorei
CC   (AmEPV)
XX
SQ   Sequence 32392 BP; 13748 A; 2577 C; 2550 G; 13517 T; 0 U; 0 Other;

  Query Match             3.7%;  Score 49;  DB 6;  Length 32392;
  Best Local Similarity   52.2%;  Pred. No. 0.85;
  Matches  109;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy       1132 TGCTTTGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATA 1191
             |||||  ||  |||| || | |  || ||  |  ||||| | ||||  | || || |
Db      22647 TTCTTTAAAAGATATAGAAAATCTTATTAATTTAGAAAAATTAGATTGTTCTAATATAAA 22706

Qy       1192 ATGGCTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTC 1251
             |      |||| | |||||    ||   ||||  ||    | |||||  |  |  ||  |
Db      22707 ACTAAATTCTTAAATATAAAACAAAATATGTTAATTTATAAATTAGATTGTTATAGTATA 22766

Qy       1252 TACATGTTAATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATC 1311
              |||| ||| | | | |  |  |     ||  ||  | | |  |     || | | ||| ||
Db      22767 AACATTTTATTTTTTATATTAAATTTGTAATTGAAATATTATAACATAAATAAAATATTC 22826

Qy       1312 TTTAAAAGATTCAAGAGTTCTTATGAGTA 1340
              | | |   ||   | |  |  |||   ||
Db      22827 TATTATGTATAATATATCTACTATTCTTA 22855


RESULT 13
ACL35887/c
ID   ACL35887 standard; cDNA; 2000 BP.
XX
AC   ACL35887;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Rice stress-regulated promoter SEQ ID NO:14450.
XX
KW   ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW   agriculture.
XX
OS   Oryza sativa.
XX
PN   WO2003008540-A2.
XX
PD   30-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US019668.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   24-AUG-2001; 2001US-0314662P.
PR   26-SEP-2001; 2001US-0325277P.
PR   21-NOV-2001; 2001US-0332132P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Kreps J,  Briggs SP,  Cooper B,  Glazebrook J,  Goff SA,  Katagiri F;
PI   Mougham T,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-248011/24.
XX
PT   New stress-responsive nucleic acid, useful for altering the
PT   responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT   stress, salt stress or osmotic stress.
XX
PS   Claim 48; SEQ ID NO 14450; 89pp; English.
XX
CC   The invention relates to novel abiotic stress responsive polynucleotides
CC   and polypeptides. Also disclosed are vectors, expression cassettes, host
CC   cells, and plants containing such polynucleotides. Also disclosed are
CC   methods for using the polynucleotides and polypeptides to alter the
CC   responsiveness of a plant to abiotic stress. The invention is useful in
CC   agriculture. The nucleic acid is useful for determining whether a test
CC   plant has been exposed to an abiotic stress condition. It is also useful
CC   for selecting an agent that alters abiotic stress regulated
CC   polynucleotide expression in a plant cell, and to identify a homolog or
CC   ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC   molecule and the polypeptide encoded by it are useful in altering the
CC   responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC   stress, osmotic stress or any of their combinations. The present sequence
CC   is used in the exemplification of the invention
XX
SQ   Sequence 2000 BP; 410 A; 329 C; 253 G; 416 T; 0 U; 592 Other;

  Query Match             3.6%;  Score 48.8;  DB 11;  Length 2000;
  Best Local Similarity   16.5%;  Pred. No. 0.5;
  Matches  126;  Conservative  274;  Mismatches  356;  Indels    8;  Gaps    2;

Qy        585 ACTATATGCTACGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATA 644
              |||| |    || | |   | | | :| || |  | || ||        |   |
Db       1021 ACTAAAATAGACATATAGTACAACGWTATCGGTTAGTACCACGAAKAATTTGACTRATAT 962

Qy        645 TAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGG 704
              |  :||  || || |||:| ::|   :|:  ::: : :  :| :  |: :: ::    :
Db        961 AACTYATCGTATAATAATWTRMYTMATMTMAMWWMYSRWMYRATRMTAWAMWTMWWCTCR 902

Qy        705 TGGTATTTCATGGGTTAATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGAT 764
              : ::::: : :: |:|::  : ::: :     :  :|:|    :||: : :  : :|
Db        901 WARYMKWASSKRTGYTWRRWWTWYWCYSRTMKWAWWAKGTGYKWTTRRYTWKYKSYGATS 842

Qy        765 GTCAGGGTATTTTATTTATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTAT 824
              : |         :|:::: |  ||  |:|      |  ::  :|: :|  : :: :::
Db        841 TWWAWATRTARCWAWKWRCTGMGTGKTKATKTAWTATAWRKAWMAMARTGCWRMKKGWWY 782

Qy        825 ATATGCAGGGCAGTTAGGGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCT 884
              |:::         : |:   :::::  | :::: : :   | ::|:| :  :  | :::
Db        781 AWWWWWTWYWRGKATWAYMMWKYMRMTSRMMYRWARKTTTCYKGWATKATMMAKWSRSKS 722

Qy        885 GGCTTGGATATCCACTAAATATCCAAAATGGGGA---GTAATGGTTTCCAACAAACAGCT 941
              :::  :   ::::  :::: || :: : :|:|  :   |:: :: : :   |:|: : :::
Db        721 MSMYGKMSGWYYSSTKWMWGATRMWGMKWGKGCRRCGGYMKWSMKSKWTCASAWGMTKYK 662

Qy        942 TGCGGAACTGAGTAAATCGTTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGAC 1001
              :::: : | |  | ::::: :: :|  |:  :: : :   :: |  ::  ::  |:::
Db        661 YSYRCRGCGGTWTTRWWMRAWMCMAKWTKTGSYYWCRGYSKYKTTWYWRAWWWSCWKKTG 602

Qy       1002 AGGATTAACTGGTGTATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGG 1061
              |: :::: |: :|| :::  :: : : :::: :|: |  :: : : |::|| ::|
Db        601 TGRTKYMWGTSTYGTTKMYSAYWKRMWSKKRWGRTKYATGWKYWGYTGWYCAYWSAAMCW 542

Qy       1062 CGAGCGGTTTTTGGGATTGCAGGATTTTTTCTTTTTATCT-----TTAGCAATTATTGGT 1116
                ::  :| |  :::  ::   : ::  :    : ::| |         |   : | :  :
Db        541 KMRKGTKTRTGRKSRGCWRTKCKTYWRMYRWMGYRWRTGTSRGRGGMAATKMSTKYRWKR 482

Qy       1117 AATCACATTGTAGCTTGCTTTGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACA 1176
              :: :| ||:::         |::     ::::  :|:|  :  |:     :|    :|: :
Db        481 WWRMAWATKSWKWRGMMWTKWCTKYAYKRWYMRAKAYGKKTTWGCSTYGKCTTWAYTRAS 422

Qy       1177 TTGGCATCATGTATAATGGCTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTA 1236
              :  :  |   :| :   :|::: :|| | |    ::| : | |:  ::  :|: ::::|
Db        421 YGAKWCTRTGKTTKGTCRGMKYRMTTAAGTKAKYYKAKRKTTTKTMWKMRYGSKWRWKTT 362

Qy       1237 GAGTACTCGAGGTTCTACATGTTAATGTATGCAGCACTAACGTGGTTATATTTTGTTCCT 1296
                |: :  ::| | ::: :: : :: : :          ||::   | ::::|   |
Db        361 AGYTRAKWKTYKYTRTTSWWRSKWKKTWWRKKWWRTKGTYKCTGKWRTMAMWWKKTARRT 302

Qy       1297 CAAACTTATATAATCTTTAAAAGATTCAAGAGTTCTTATGAGTA 1340
              :::  :: |    : :: |::|| :  :   :::::  :  :|:
Db        301 SWWKTKKYTTMKGKRWKCAMRAGTKAKWTCTRWWYYMYWAKRTR 258


RESULT 14
ADA71938
ID   ADA71938 standard; DNA; 2000 BP.
```

XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 3.6%; Score 48.6; DB 8; Length 2000;
Best Local Similarity 13.6%; Pred. No. 0.56;
Matches 41; Conservative 145; Mismatches 114; Indels 1; Gaps 1;

Qy 37 GCAATTCGTTATGTTATTTTTGGTAAGATAATATCTGTATTAACGGGTTTACTGTTAATA 96
Db 717 RMAMTGMKYSGRYWTSWYKYCKCSWKYRSMWYYWSWWWAKTWKMWRRYATRMMWMWYRYS 776

Qy 97 ATGTTAATATCACACCATTTATCTAAAGACGCACAGGGCTATTATTATACATTTAATTCA 156
Db 777 MKWYTWCTMWGYWWYWWRTYMKMRYMWYKCTKTYWYWSATYWTGTWAAWWMAKTKMRMGM 836

Qy 157 GTAGTGGCACTACAGATAATA-TTTGAATTGGGGCTATCAACGGTAATCATTCAATTCGC 215
Db 837 TGAKTRGRARKARYWWKWATWCATKRWMTKGKGAKWAWTWMAKAWRKYYWSWMRAWYYYY 896

Qy 216 TAGCCATGAAATGTCAGCGTTAAAATATGATTATTCTGAACGAGATATTATAGGTGAAAG 275
Db 897 KTRRTRYKTCWWKARWGSWAYWRMWWKGSAKMWWMWKGGRWGWTKYWYWYCTTWKMACGR 956

Qy 276 TAAAAATAAGCAACGTTACCTATCGTTATTTCGGTTGGCAATAAAATGGTATGCAGTAAT 335
Db 957 ATKYMCCAGWWAMYSYSWTRTYWMRTWRWMWASSRTAKRMARMMWKTRAWSKSYARAYWK 1016

Qy 336 A 336
:

Db 1017 M 1017

RESULT 15
ADZ71022
ID ADZ71022 standard; DNA; 1162 BP.
XX
AC ADZ71022;
XX
DT 14-JUL-2005 (first entry)
XX
DE Human chromosome 2 contig DNA SEQ ID NO 146.
XX
KW ds; matrix attachment region; MAR; protein production.
XX
OS Homo sapiens.
XX
PN WO2005040377-A2.
XX
PD 06-MAY-2005.
XX
PF 22-OCT-2004; 2004WO-EP011974.
XX
PR 24-OCT-2003; 2003US-0513574P.
PR 06-FEB-2004; 2004EP-00002722.
XX
PA (SELE-) SELEXIS SA.
XX
PI Mermod N, Girod PA, Bucher P, Nguyen D, Calabrese D, Saugy D;
PI Puttini S;
XX
DR WPI; 2005-333507/34.
XX
PT New purified and isolated DNA sequence having protein production
PT increasing activity comprises a bent DNA element and a binding site for a
PT DNA binding protein, useful for increasing protein production activity in
PT eukaryotic host cell.
XX
PS Disclosure; SEQ ID NO 146; 282pp; English.
XX
CC The invention relates to a purified and isolated DNA sequence having
CC protein production increasing activity comprising at least one bent DNA
CC element, and at least one binding site for a DNA binding protein. The
CC purified and isolated DNA sequence comprising a first and a second
CC isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC MAR nucleotide sequence selected from a purified and isolated DNA
CC sequence above, a purified and isolated MAR DNA above, a purified and
CC isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC sequence complementary to it, its molecular chimera, or its combinations
CC and variants, is useful for increasing protein production activity in a
CC eukaryotic host cell. The present sequence represents a human chromosome
CC 2 contig DNA.
XX
SQ Sequence 1162 BP; 486 A; 64 C; 6 G; 606 T; 0 U; 0 Other;

Query Match 3.6%; Score 48; DB 14; Length 1162;
Best Local Similarity 46.6%; Pred. No. 0.68;
Matches 153; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

Qy 1007 TAACTGGTGTATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGC 1066
Db 835 TATCTATTATATATATTATATATATACTATCTATTATATCTATTATATATATTATATATA 894

Qy 1067 GGTTTTTGGGATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTG 1126
Db 895 TACTATCTATTATATATCTATTATATATATTATTTTATATTATATATAGTATCTATTACA 954

Qy 1127 TAGCTTGCTTTGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCAT 1186
Db 955 TATATTATATTATATTATATATAATATCTATTATATATATTATATTATATTATAAATAAT 1014

Qy 1187 GTATAATGGCTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGA 1246

```
Db       1015 ATATATAATATCTGTTATATATAATAGATATTATATATAATATATAATATATATAATAGA 1074

Qy       1247 GGTTCTACATGTTAATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATA 1306
                || || || |||   |||  |  |   |      | ||||  |     | |||  ||||
Db       1075 TATTATATATATTATATTATATAATATATAATATATAATATAATTAATATAAAATATATA 1134

Qy       1307 TAATCTTTAAAAGATTCAAGAGTTCTTA 1334
              |||| | |||   ||  || |  |  ||
Db       1135 TAATATATAATTAATATAATATGTAATA 1162
```

Search completed: May 11, 2006, 09:42:53
Job time : 614.03 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 09:43:19 ; Search time 4178.12 Seconds
(without alignments)
15016.694 Million cell updates/sec

Title: US-09-423-093B-3_COPY_1019_2359
Perfect score: 1341
Sequence: 1 atgaatcgtattattagaat.....tcaagagttcttatgagtaa 1341

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	75	5.6	1101	10	CNS0039G	AL063921 Drosophil
	2	58.8	4.4	948	10	CNS0159T	AL105179 Drosophil
c	3	58	4.3	895	11	CNS06FSV	AL396821 T7 end of
	4	54.2	4.0	1092	10	CNS020K7	AL175696 Tetraodon
	5	54.2	4.0	1101	10	CNS000D1	AL065414 Drosophil
	6	52.4	3.9	859	10	CNS00KLL	AL077728 Drosophil
c	7	52.4	3.9	1101	10	CNS006DU	AL065139 Drosophil
	8	52.4	3.9	1225	10	CNS0161D	AL106171 Drosophil
	9	52.2	3.9	902	10	CZ522268	CZ522268 GMW2-38B1
c	10	52.2	3.9	1101	10	CNS00EVL	AL069706 Drosophil
	11	52	3.9	743	10	CE685136	CE685136 tigr-gss-
c	12	51.6	3.8	914	10	CNS00ZJY	AL097768 Drosophil
	13	51.4	3.8	894	7	CN757547	CN757547 ID0AAA1DE
	14	51	3.8	1101	10	CNS000B8	AL063632 Drosophil
c	15	50.8	3.8	734	10	CNS010MP	AL099163 Drosophil
	16	50.8	3.8	780	10	BX150770	BX150770 Danio rer
	17	50.6	3.8	1147	1	AJ925698	AJ925698 AJ925698
	18	50.6	3.8	1310	10	AG396658	AG396658 Mus muscu
	19	50.4	3.8	891	10	CZ974720	CZ974720 175677 To
	20	50	3.7	1101	10	CNS00LT2	AL078714 Drosophil
	21	49.6	3.7	1086	10	CNS00YXK	AL096962 Drosophil
	22	49.6	3.7	1101	10	CNS00Z2H	AL097139 Drosophil
c	23	49.4	3.7	1492	1	AJ928220	AJ928220 AJ928220
c	24	49.2	3.7	964	10	CNS006N9	AL065781 Drosophil
c	25	49	3.7	729	11	CR135346	CR135346 Reverse s
c	26	49	3.7	730	10	BX966525	BX966525 Reverse s
c	27	49	3.7	866	11	CR177886	CR177886 Reverse s
	28	49	3.7	1101	10	CNS0106X	AL098595 Drosophil
	29	48.8	3.6	608	7	CV740444	CV740444 SJAL_027_
c	30	48.8	3.6	878	10	CNS0187R	AL108993 Drosophil
	31	48.4	3.6	928	10	CNS00DKY	AL071865 Drosophil
c	32	48.4	3.6	1101	10	CNS012I9	AL101595 Drosophil
c	33	48.4	3.6	1303	9	BZ555610	BZ555610 pacs1-60_
	34	48.2	3.6	518	9	BH584236	BH584236 BOHNY69TR
	35	48.2	3.6	994	10	CNS015XG	AL106030 Drosophil
c	36	48	3.6	479	9	AZ074878	AZ074878 RPCI-23-4
c	37	48	3.6	495	7	CO876873	CO876873 BovGen_05
c	38	48	3.6	843	10	CNS00CS1	AL059666 Drosophil
c	39	48	3.6	970	10	CNS0182A	AL108796 Drosophil
	40	47.8	3.6	585	10	CE667331	CE667331 tigr-gss-
c	41	47.6	3.5	994	11	CNS04NOJ	AL298972 Tetraodon
c	42	47.6	3.5	1493	10	CL078589	CL078589 CH216-151
	43	47.4	3.5	364	1	AU038586	AU038586 AU038586
c	44	47.2	3.5	573	8	CV946518	CV946518 PV020C5 z
	45	47.2	3.5	661	10	CNS02QVJ	AL209800 Tetraodon

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

```
  Query Match             5.6%;  Score 75;  DB 10;  Length 1101;
  Best Local Similarity   19.9%;  Pred. No. 6.6e-06;
  Matches  137;  Conservative  272;  Mismatches  274;  Indels    6;  Gaps    1;

Qy        127 GCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTTGAATTG 186
              | | : |: |:: :|  :: ::|:::|:  : |:::: | :| : |::: | |:  ||:
Db       1101 GKARRWGDDTWDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWW 1042

Qy        187 GGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAATATGAT 246
                ::  ::   : :: : | | |:   |:   |::|:: |  :| |: ::|  : :
Db       1041 WWWWWWATWDTWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADG 982

Qy        247 TATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGTTATTT 306
                |:  ::: :::: :::::|  ::|| ||::|  :: :::: ::: :  : :
Db        981 AGRRDGGRKRKDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRK 922

Qy        307 CGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTCGGTCCC 366
              ::::|  : :::   :: ::: :|| :::: :::: ::    |: :| :|    ::
Db        921 ADDDDGAGDKDDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWW 862

Qy        367 ATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGG 426
              : : :::| :   ::::| : ::|:|  ::: : :|    :  ::|: | ::|| ::::
Db        861 GWADADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRR 802

Qy        427 TTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCT 486
              :   ::::: |::::: | :  :|||   :|:::: :|:|:|  :  :  |:: ::: : :
Db        801 DRKRADDKRDAADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRD 742

Qy        487 GAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCT 546
              :|   |::::: ::::::: : :::: :|:: |::::::: ::   ||: : || :::
Db        741 RAGTAGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRR 682

Qy        547 GGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATA 606
              ::  :: |:  : :||  ||  |: :::::::  :| :  :  |  | :: |  |:::
Db        681 DRATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDW 622

Qy        607 GCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTC 666
              :::|  |:: : :|   : :  :  ::: :   : | :::      |:||:|:  :|
Db        621 AKWDWKTRADRWDRWAADTWTDARKADRDWAKARAWRARRDR------ARAARADRRWTT 568

Qy        667 CTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAA 726
              :    :||:||: |::|::|:|  |||: | :  :  | |||  |   ||||: :|
Db        567 KGKTTTATWTTWAARAAWWAWWAWATTTATWTTTWTTTWTTTTTTTTTTTWTTTAAWWWA 508

Qy        727 ATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTT 786
              |::| |   :|  |:     ||:   | |   ||   : : |   :  :|: |||:||||
Db        507 AWWTATWAAWTAAAWAAAAAAAWAATTTTTTTTTTTTWAWTAAAWWTAWTWTTTTWTTTT 448

Qy        787 GTTATGACCCCCATTGCATTCAAATATTT 815
              : ||       :||  | | |  | :||
Db        447 WWAATTTTTTWTWTTWWAATTATTTTWTT 419
```

```
RESULT 2
CNS0159T
LOCUS       CNS0159T                948 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN13A12 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL105179
VERSION     AL105179.1  GI:5617193
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 948)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .948
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN13A12"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN
```

```
  Query Match             4.4%;  Score 58.8;  DB 10;  Length 948;
  Best Local Similarity   18.7%;  Pred. No. 0.025;
  Matches   62;  Conservative  135;  Mismatches  134;  Indels    0;  Gaps    0;

Qy        398 GCTTAGGTGTACCTTGGCAAGGGGCATGGTTATTATTAACAATAGTTACAGCTTTTAATA 457
              ::: ::  :  : | | :::    ||   :||:::: :::|   |:::|:   |   :
Db        575 KMKNMKMMKNTMTTMGHMHMTHTMTKTGMMMMTTHHHMGMMMTTTMTMMMGMVMMTTTMR 634

Qy        458 TTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATA 517
              || |:   :| :      ||:  ::  :| ::| |:    : ::   :      |:|:: :
Db        635 MTTMTMMMTHTMMMHMTTTTKGMVKGMMTMVHGMGMMMMMKVKMKGGVGNNGMGDGMMMM 694

Qy        518 AAATGAGAATGTATCAGTCGCTGTTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTG 577
              ::  ::: : : :::    :  :  :::  : ::  ::: :::::::::::: :: :: ::
Db        695 VMKGRMKGMMKGMKMKMMKKKMKGVMKVGMVVMKGMKKKKMKKMMKMHKMHNHMMKMTKK 754

Qy        578 GCTTTGGACTATATGCTACGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATAT 637
              ::|::::   :: :: ::  : :  || |      ::  : :| : : | :|| |  | ||
Db        755 KMTKKKKNKNMWKKKAKDKAKKKAWAAAAAAKAKAKKAKKAKAKKKKDAAWTTTTTAAAAT 814

Qy        638 TTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATA 697
              |:| :|| |  : ||||::|  || :|:  :  :|| :| ::|  |:| ||| ||:   |
Db        815 TWTAWTAAATTWTTTTTWWATTAAAWTKTAWWAWATWWTAWWATAAKATAAAAAAWATAA 874

Qy        698 CTGAAGGTGGTATTTCATGGGTTAATGAAAT 728
              ::|:::::: : ::  :: :::|:::|:|:
Db        875 AKDAMKKKKKAMMKKKKHKMKKKAMKKAMAH 905
```

```
RESULT 3
CNS06FSV/c
LOCUS       CNS06FSV                895 bp    DNA     linear   GSS 17-JUN-2001
DEFINITION  T7 end of clone AR0AA032H05 of library AR0AA from strain CBS 732 of
            Zygosaccharomyces rouxii, genomic survey sequence.
ACCESSION   AL396821
VERSION     AL396821.1  GI:12149222
KEYWORDS    GSS.
SOURCE      Zygosaccharomyces rouxii
  ORGANISM  Zygosaccharomyces rouxii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.
REFERENCE   1  (bases 1 to 895)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   2  (bases 1 to 895)
  AUTHORS   de Montigny,J., Straub,M., Potier,S., Tekaia,F., Dujon,B.,
```

```
          Wincker,P., Artiguenave,F. and Souciet,J.
  TITLE   Genomic exploration of the hemiascomycetous yeasts: 8.
          Zygosaccharomyces rouxii
  JOURNAL FEBS Lett. 487 (1), 52-55 (2000)
   PUBMED 11152883
REFERENCE 3  (bases 1 to 895)
  AUTHORS Genoscope.
  TITLE   Direct Submission
  JOURNAL Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
          2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
          seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT   This GSS is part of a random genomic sequencing program of thirteen
          yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
          exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
          Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
          lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
          angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
          Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
          5 kb were prepared and both extremities were sequenced. See
          keywords for description of this sequence and for the sequence of
          the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .895
                     /organism="Zygosaccharomyces rouxii"
                     /mol_type="genomic DNA"
                     /strain="CBS 732"
                     /db_xref="taxon:4956"
                     /clone="AR0AA032H05"
                     /clone_lib="AR0AA"
                     /note="end : T7"
     tRNA            <19. .>93
                     /product="tRNA-Thr"
                     /note="putative mitochondrial tRNA_thr gene"
                     /evidence=not_experimental
ORIGIN

  Query Match             4.3%;  Score 58;  DB 11;  Length 895;
  Best Local Similarity   41.0%;  Pred. No. 0.037;
  Matches  181;  Conservative   18;  Mismatches  243;  Indels    0;  Gaps    0;

Qy        431 TATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAG 490
              ||| | | |  ||      ||||  | ||||| || ||| | |  ||| | |   |
Db        885 TATATATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 826

Qy        491 GGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTA 550
                  |   ||  ||  | :| |   | ||  |   | |||   |   |  |   |  ||
Db        825 TTTWTTTTTTTTTTTTTTTWTTTTTNNATAAATATTTTTTATATATATTTTATTTTTTATA 766

Qy        551 TATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTA 610
              || |   |  :   ||  |   || |   |||    |||||  :|    |   |||  ||
Db        765 TANTTATATAWTTTTTTNTNTATATTTTTTTTTTTTTATATATWANAAWTTATATATATA 706

Qy        611 TTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGC 670
              | | |   |      ||::| |  | ||:   || ||  | | | || | |  | |  |
Db        705 TATNATATATATATATWWTATNTAAATWAATATNTATATTTATAAATATATATATATTAT 646

Qy        671 AATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATAT 730
              | |  |  |   ||| |: |:||| || : |:        |  :|    |  :    |||
Db        645 ANTTATATATTTATATAWTTWAATTTATWWAWATANTATNTNWWTATWNTNTWATTTTAT 586

Qy        731 TTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTA 790
              :|       |||      ||   | ||  |   || |     | |  ::||| || | |||
Db        585 WTATATATAAATATATTATATWTATANTTAATATATATWATTTTAAWWTTTNTTATTTTA 526

Qy        791 TGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTT 850
                :     |||  ||  ||     ||       |||||||  |    |  ||   || | |
Db        525 ATWATTATATTATATATAATATATTTWTTWATATATATTTATATNATATATTTATTTATN 466

Qy        851 TAACATTATGCAATATGGTAAT 872
              :|  ||:|  | |    | ||
Db        465 AWATNTTWTATATTTATATTAT 444
```

```
RESULT 4
CNS020K7
LOCUS       CNS020K7                1092 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            222L11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL175696
VERSION     AL175696.1  GI:7813753
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1092)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="222L11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG222CF06LP1
                     end : T7"
ORIGIN

  Query Match             4.0%;  Score 54.2;  DB 10;  Length 1092;
  Best Local Similarity   35.2%;  Pred. No. 0.26;
  Matches  158;  Conservative   58;  Mismatches  233;  Indels    0;  Gaps    0;

Qy         10 ATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGTAAGATAATA 69
              | : ||| : |   :  :  | | |||  ||::  || | | :        :|| ::::::
Db        613 AAWTTTATWTTTAAWWWWAAAAAAAAAAAAAAWWTTTTTTTTAWAAAAAAAAWAAWWWWWW 672

Qy         70 TCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCTAAAGACGCA 129
              : |  | ||    |||  | ||  |  | || | |  | |  |   |  : :|     |
Db        673 WTTAAAAAAAAAATTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAWTWAATWAAA 732

Qy        130 CAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTTGAATTGGGG 189
                :     |: |: |:: |:||: ||   :| :   :  | | | |:::|:| |:||
Db        733 WWAAAAAAWAAWAAWWWAWTTWTTTTTTWAWWWWWWWAAAAWAAAWWWTWTTAWTTTTAW 792

Qy        190 CTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTAT 249
              :|| |:   : ||  |  |:      :   |:  |:|   |   : || |  |  |||
Db        793 WWATWAWWTTWTATTTTAAAWAAAANAWTTTAWTTAWTAAAAAAWWAAATAATTWTTTAA 852

Qy        250 TCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCGG 309
```

```
Db        853 AAWAWTTTAWAAATTWWAATWTAAAAAAAAAAATTAAAAAAWAAWTWWTTTTTTTTWTWA 912

Qy        310 TTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATC 369

Db        913 WWTWTAAATWTTTTTTTTTTTTTWWTTTTWTAAWAAAAAAAAAAAAAATATAAAWNAAAA 972

Qy        370 GGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGGTTA 429

Db        973 WWTTAATTTATTTTTAAAAAAAAAAAAAAAATAAAATATAAAWAAATAAAATATTTATATA 1032

Qy        430 TTATTAACAATAGTTACAGCTTTTAATAT 458

Db       1033 TTAAAAAAAAAAAAAATAAATTAAAAAWT 1061


RESULT 5
CNS000D1
LOCUS       CNS000D1                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR01J16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065414
VERSION     AL065414.1  GI:4938827
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR01J16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             4.0%;  Score 54.2;  DB 10;  Length 1101;
  Best Local Similarity   15.3%;  Pred. No. 0.26;
  Matches   82;  Conservative  235;  Mismatches  215;  Indels    3;  Gaps    1;

Qy        622 ATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAAG 681

Db        559 AWYTTTTTTWTATTTWKAANAAAAAWAAAWAAAAAAAAAAWAGARGATNDDKBAKDWKAAA 618

Qy        682 CATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCTATGCAA 741

Db        619 DAKAKAAAWHAGTWWTKKDTKKKKKKKRKWGKAKRAATDDTAWTATTWKKTAKKDAGKAA 678
```

```
Qy        742 TGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTT---TGTTATGACCCCC 798

Db        679 KWKADRTKAWKGAKGAGARGKATAKARAKKRWKDDKDTKKKKTKTTATKKKTTTTKRAAT 738

Qy        799 ATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTTTAACATTA 858

Db        739 DKWKWRKKKATAGAWDWDDARRGRGGRWGDKRRRTGARGDKDKTGKKKKTGDKRDKGGGD 798

Qy        859 TGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACTAAATATCCAAAATGGGGA 918

Db        799 GDRGDGGAGGGDRDRDARARRARWDTKDTRGAKDKADKKKAGRKRKWWTRKTWKRDTADW 858

Qy        919 GTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCGTTCAAAAGTGCAGTAATG 978

Db        859 KAWAAWRWRWRRDRKDRRGRRDRWRRWRGADKKWKWDDDDTAKDDWTKKDTDWTAGATAG 918

Qy        979 CAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATACATTTCATTATGGTTATTG 1038

Db        919 WDDDKAKDGDWKRGTDWRTWKRTWKGWGRKGDKGRKTAGRAGDGKDKKKGKDKDRTDDAK 978

Qy       1039 AAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTGCAGGATTTTTTCTTTTTA 1098

Db        979 ATDGKDTWRDTTDTAKRRRRKRRRRAGRGRDRKGDKKKRWRARWRKGKKKKTKDKDDDDG 1038

Qy       1099 TCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCAACCTATATAAGAG 1153

Db       1039 KRWGRAKAKADAAKAKDDADAADAAAWAKAKAKDKDKAKKKADADDRKDKKAKAR 1093


RESULT 6
CNS00KLL
LOCUS       CNS00KLL                859 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR17N04 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL077728
VERSION     AL077728.1  GI:4957306
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 859)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .859
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR17N04"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN
```

```
  Query Match             3.9%;  Score 52.4;  DB 10;  Length 859;
  Best Local Similarity   23.2%;  Pred. No. 0.64;
  Matches   85;  Conservative  123;  Mismatches  158;  Indels    0;  Gaps    0;

Qy        974 TAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATACATTTCATTATGGT 1033
              |  |        |:|:::|:| | ||:     |       |::    :   ||:  :| | :|
Db        459 TNNTTTNNNNNTYCYBYTKTTTGTTKTTGNGTGNGTKTKKGTGTKKKKTTWTGWTTTNKT 518

Qy       1034 TATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTGCAGGATTTTTTCT 1093
              | ||    || | |  ||      ::|| : | :|  |::: |:: | :|: |:||:| |
Db        519 TTTTTTTTTTTTTTTTTTTTKKKKKGGGKTGGKGGATKWKTGWKAGARGRGTKTTKTTT 578

Qy       1094 TTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCAACCTATATAAGAG 1153
              ||||||  ||     |  ||:  :|: :     || ::| : : ::|: |     : ::
Db        579 TTTTATKTTRTTTAKATADATKTRTDTTTTTTTKWGDDARTKKTDDATAWRRGWDGTTTT 638

Qy       1154 CGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTCTTGACTATAACTA 1213
                : :|||   |||||  |    :||   :| |  :| :|: : : | :  |  | :
Db        639 WAAWAWAAAAAAAAAAATTTTWWTTAWWWTWAAAWAAWAWADKWWTWWAAAAAARWKGGG 698

Qy       1214 CAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACATGTTAATGTATGCAGCAC 1273
                ||    ::::::::|      ::::::: : ::: :  ::   :  :|::: :: :| :
Db        699 RAAAARRRDDKKKDGGGRKRGDRRDRWRRKDRRRADGAKRDKGDGGDADRWGDRDDGARR 758

Qy       1274 TAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTTAAAAGATTCAAGAGTTCTT 1333
              ::: :::::::::::: :::    ::  :::::: :: : : |:::::| :||:::: :|
Db        759 WWDKDDDDDDDDRKDKRKDDGWRKDRKDDDWDDWGWKRWGWKADRDWKTDDAGDDWDKDT 818

Qy       1334 ATGAGT 1339
              ::  ||
Db        819 DDWKGT 824


RESULT 7
CNS006DU/c
LOCUS       CNS006DU                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR13D04 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065139
VERSION     AL065139.1  GI:4944309
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR13D04"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             3.9%;  Score 52.4;  DB 10;  Length 1101;
  Best Local Similarity   33.1%;  Pred. No. 0.66;
  Matches  177;  Conservative   78;  Mismatches  279;  Indels    1;  Gaps    1;

Qy        245 ATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGTTAT 304
              ||:||: : ||  | ::| ||||  |  | : |: |   |:    ::::  : : ::: :
Db       1006 ATWATWGWAAAWAATWWAATATAWATATANRAARTATATAKWKKKKDKDKKKKKKKKKKK 947

Qy        305 TTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTA-ATAATATTAATAGTCGGT 363
              :: :::::   :    :::: ::  ::  : : ::::  | |       :   : |  |
Db        946 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKRKAAGAGNNNNNKNNNRNTTTTT 887

Qy        364 CCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGCA 423
                :|     |:  ||:|   :    ||||  |||   ||| :      ||  |:||  |
Db        886 TTTWTWTTAAAWTATTKTAGARATAKAAAAAGAGGGGAGGGKGGGGRGGGWGARGGATA 827

Qy        424 TGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTC 483
              ||  :||   ::|       : | |  | |:: |||:|||  | | | |:|| ||
Db        826 TGTAWATATAKRAKGTGTAWAAWATWTATWDTTATWTTTTGTTTGTATRTATTTAWATAW 767

Qy        484 GCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTA 543
                :   |||||  |||       |   |     :: ||: | :   : |   :       :
Db        766 AADAKKGGGAGGGGGGGGGRAAAAAAAAAAARRRAAWTTAWAGGKATTAAWAATWTATGDK 707

Qy        544 GCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCA 603
              |:|| |||: |   ||    || :| | | |   |           |     |     |
Db        706 TTTKGTGTATKWGTGATATAWATAAKTTTAATTANTATGTTGAATATTATATTGAWATAA 647

Qy        604 ATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATT 663
              |||  | ||  |     ||| ||  |    || || | |  | | ||||  | ||||||
Db        646 ATATTTKTTGAATAATATATAATTATGATAATTTTATAAATTTATTATTAAATAAAAATA 587

Qy        664 TTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAAT 723
              |   ||     ||| ||  |  ||||| || ||: |: |  |    ::  ||     |:
Db        586 TAATTGTTTAATTTTAATAAATAAAATTAAAATRTTRTATTTATATKKATATTTAAAAWA 527

Qy        724 GAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTA 778
              : | |||   ||||  |||| |    | :   |:: ||  |   |||| :   |
Db        526 TWTAAATTADTTATGTNATGGTGTGNAGGKAAKAKKGGGGGGGGGGGGGTKKNNAA 472


RESULT 8
CNS0161D
LOCUS       CNS0161D                1225 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106171
VERSION     AL106171.1  GI:5620504
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1225)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
```

```
          and Genevieve Payan. It has been constructed in the vector
          pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1225
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15C18"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             3.9%;  Score 52.4;  DB 10;  Length 1225;
  Best Local Similarity   26.7%;  Pred. No. 0.66;
  Matches  168;  Conservative  137;  Mismatches  324;  Indels    0;  Gaps    0;

Qy         32 ATAAAGCAATTCGTTATGTTATTTTTGGTAAGATAATATCTGTATTAACGGGTTTACTGT 91
Db        576 WTWATTTTWTTWWTTTTTTTATTTTTWTYATYAATAAWTWTTTTWTTTAAYYMCCKYTTW 635

Qy         92 TAATAATGTTAATATCACACCATTTATCTAAAGACGCACAGGGCTATTATTATACATTTA 151
Db        636 TAAAAAAATTAACCTCCMYYTHYYAWTAYMWTCNTTMTWHYYTTCCYTCCYYCCCTCCCC 695

Qy        152 ATTCAGTAGTGGCACTACAGATAATATTTGAATTGGGGCTATCAACGGTAATCATTCAAT 211
Db        696 CTCTCYTTYTYTTTYTHTCTTTTTTTTTYTHAHAWCCKYYYCHMMAAMAMAYAKCCKBKK 755

Qy        212 TCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTATTCTGAACGAGATATTATAGGTG 271
Db        756 KNNKWCKMKMKHMAADAAKAMAMNWAAHHWTTTYTWAAHAYNAAAAATTTAAAATCKCAA 815

Qy        272 AAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCGGTTGGCAATAAAATGGTATGCAG 331
Db        816 AMAMAAMAAAMMAAAAACCCMCCVCCKKKKDKKKKKKKKDDKDAAKRARADDRKAAAAAA 875

Qy        332 TAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGTATGTTTTTTTTACGCAAA 391
Db        876 AAATATAAAAAAAAAAAAAAAAAWWWATTYTYYYAYAWAAAYAWATAAWWAWTTWTTAAAW 935

Qy        392 AAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGGTTATTATTAACAATAGTTACAGCTT 451
Db        936 WAAAAAAAAAWAWWTTTYAAAAAAWTWTWAAATAAWTTATTTTTTTTTAWWWTTTATWWA 995

Qy        452 TTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGGTTAATTACTGATG 511
Db        996 TAATTTATWTTTWTTTTTTTTTWTAAAAWAAAAAAAAATATATTTTWTWAWWWTWTWATTT 1055

Qy        512 TGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATATTGGCAGTAAGCTTACTTA 571
Db       1056 TTATATATWTAYTTTTATATTMWAWAWWATWWTTWTWTWTTWTTWWWTWWWTAATATTTT 1115

Qy        572 TTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTTCAGGGACTATCATATTCT 631
Db       1116 TWTTTTATWTAWAATWATATAAAAAWTWTTTWWWATAATAWATWTATATATTTTTATAAA 1175

Qy        632 CCATATTTTCATATAAGTATTTTAAAAAA 660
Db       1176 WAAWWTTTTWTWTWWWWTWTWTWWAAAAA 1204


RESULT 9
CZ522268
LOCUS       CZ522268                 902 bp    DNA     linear   GSS 11-MAY-2005
DEFINITION  GMW2-38B19a.g2 GMW2 Glycine max genomic, genomic survey sequence.
ACCESSION   CZ522268
VERSION     CZ522268.1  GI:63247689
KEYWORDS    GSS.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 902)
  AUTHORS   Nunberg,A., Bedell,J.A., Citek,R.W., Robbins,D., McMenamy,J.,
            Peterson,S., Jones,J., Fries,J., Budiman,M.A., Nguyen,H.,
            Clifton,S.W. and Stacey,G.
  TITLE     BAC end sequences from Glycine max Williams 82 cultivar genome
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .902
                     /organism="Glycine max"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3847"
                     /clone_lib="GMW2"
                     /note="Glycine max BAC library"
ORIGIN

  Query Match             3.9%;  Score 52.2;  DB 10;  Length 902;
  Best Local Similarity   47.9%;  Pred. No. 0.71;
  Matches  175;  Conservative    0;  Mismatches  187;  Indels    3;  Gaps    1;

Qy        427 TTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCT 486
Db        217 TAAATATTAATAATTATATATTATTTTATTTTTTTTTTATTTTTTTATATTATTTTTTTT 276

Qy        487 GAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCT 546
Db        277 AATATATTTTTTTTTTTAATATTTATAAATATTATAAATATTTATAATTTTATANTATTT 336

Qy        547 GGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATA 606
Db        337 TTTTTTTT---AATATATTTTTTTTTTANTTAATTTTATTTATATATTTTTTAATTTATA 393

Qy        607 GCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTC 666
Db        394 ATTATTTTATATTTTATTTTTTTATTATNTTAATTTTTATATTAAAATTTTATTATATTTT 453

Qy        667 CTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAA 726
Db        454 ATTAATTATTTTAAAAAAAATTAATAATATTTATTATATAATAAATAATATTTTAAAATT 513

Qy        727 ATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTT 786
Db        514 AAATATATTNATTAAAATATAATATTATTTAATTTTATATATATTTTTTTTAATTAAATT 573

Qy        787 GTTAT 791
Db        574 ATAAT 578


RESULT 10
CNS00EVL/c
LOCUS       CNS00EVL                 1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
```

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"
ORIGIN

Query Match 3.9%; Score 52.2; DB 10; Length 1101;
Best Local Similarity 31.8%; Pred. No. 0.73;
Matches 198; Conservative 91; Mismatches 334; Indels 0; Gaps 0;

```
Qy        235 TTAAAATATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTAC 294
              |: :  |:|  ||::| |     | |||| |:   | ::  || | :||:   |   ||
Db       1095 TWTWTTTWTTTTTWWTATWTTAWATATATAAWTAATTWWTWTATATWTAWWWTATWATAT 1036

Qy        295 CTATCGTTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTA 354
               |   ||:|    |     | ||   |||  | ||::|  | |    |:||:  ||
Db       1035 ATTAWTTTWTAAAATATAWWTTAATAAATATATWTAWTAWWATATATAWATAWAAWTATA 976

Qy        355 ATAGTCGGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGG 414
              : : :   |    :    :|: ||  |  :|    :|::  :   ||   | |   ||
Db        975 WAWTWWAATWAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTTTTTWT 916

Qy        415 CAAGGGGCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTA 474
                :     :|   |||::|::| ||:| : |     |||  |:|| |: ||  |  | :|
Db        915 TTWTATWTWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATATTAWA 856

Qy        475 CTTTCTGTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAG 534
               ||| | |   | ::  | :   :|| :|: : |  :  : ||   :| |:| :
Db        855 TTTTTTWTWTATTWWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTWTWTTWT 796

Qy        535 TCGCTGTTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCT 594
                | |:      : :: |     :||   ||  |::   |   ::|    : |:|  |
Db        795 ATTWTATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTTWTTTT 736

Qy        595 ACGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTT 654
              |  : |  ||:   |:| | |     :||  | || |   | |||: || |:|   | ||
Db        735 ATTWTTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTWTATTTWAAATTATT 676

Qy        655 AAAAAAATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCA 714
               ||  |:::       || |   ||  |   |||| | || |       |   |    |
Db        675 TAATTAWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAAAAATA 616

Qy        715 TGGGTTAATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTAT 774
              |   || ||   |  : :  |    :||    |  :  | ||| ||     |     ::|
Db        615 TTTTTTTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTATTWWT 556

Qy        775 TTTATTTATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGG 834
              || | ||||:|| || |       |::  |   || :|:::       ::::|
Db        555 TTAAATTATWTTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWTWTTTTT 496

Qy        835 CAGTTAGGGATGTCTTTAACATT 857
               | |:|   :| | |||:: : |
```

Db 495 WATTWATAAWTTTWTTTWWAWAT 473

RESULT 11
CE685136
LOCUS CE685136 743 bp DNA linear GSS 29-SEP-2003
DEFINITION tigr-gss-dog-17000329583166 Dog Library Canis familiaris genomic, genomic survey sequence.
ACCESSION CE685136
VERSION CE685136.1 GI:37004186
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 743)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .743
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site_1: BstXI; Libraries were prepared from peripheral blood"
ORIGIN

Query Match 3.9%; Score 52; DB 10; Length 743;
Best Local Similarity 50.4%; Pred. No. 0.77;
Matches 127; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```
Qy         32 ATAAAGCAATTCGTTATGTTATTTTTGGTAAGATAATATCTGTATTAACGGGTTTACTGT 91
              |||    ||||  |||   || ||||    | ||||||     |  |     | ||
Db        123 ATATGTGAATTTTTTAAAATAATTTTTTATAAATAATAAAATAAAAAGATTATATAAATA 182

Qy         92 TAATAATGTTAATATCACACCATTTATCTAAAGACGCACAGGGCTATTATTATACATTTA 151
              ||  ||| || |||| | |  | | || |||| |   | |      ||| |||| |  ||
Db        183 TATAAATATTTATATTATATAAATGATATAAATAAATATATAATATTTAATATATAAATA 242

Qy        152 ATTCAGTAGTGGCACTACAGATAATATTTGAATTGGGGCTATCAACGGTAATCATTCAAT 211
              ||  |  | |   |  |   |  ||||    |||     ||| ||   |||| | | |||
Db        243 ATAAATAAATAATATGAATAAATATATAAATATTTAATATATAAATAATAATAAATAAAT 302

Qy        212 TCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTATTCTGAACGAGATATTATAGGTG 271
                 |    |   |    ||   |||| |||| | || |   ||  | ||| |||   |
Db        303 AATATGAATAAATATATACATATTTAATATATAAATAATAATAAATAAATAATATGAATA 362

Qy        272 AAAGTAAAAATA 283
              ||  || |||||
Db        363 AATATATAAATA 374
```

RESULT 12
CNS00ZJY/c
LOCUS CNS00ZJY 914 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN02L18 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

```
ACCESSION   AL097768
VERSION     AL097768.1  GI:5609379
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 914)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .914
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN02L18"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             3.8%;  Score 51.6;  DB 10;  Length 914;
  Best Local Similarity   29.2%;  Pred. No. 0.97;
  Matches  114;  Conservative   82;  Mismatches  193;  Indels    1;  Gaps    1;

Qy        401 TAGGTGTACCTTGGCAAGGGGCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTT 460
              |::: : :   |:  :   :: :|: || ||| |:|  ::  | | |  |||| |::|:|
Db        900 TWRRARRRARRTRTWRTTARRRWTRTTTTTTARTRATTRWTTTAARAARTTTTTADWTKT 841

Qy        461 TTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAA 520
              || :| ||| | |:   | : | :  | : :|     |  :|::|   |   ::| :::: 
Db        840 TTRKTTTTTKTTTRKRGTAKTTTRKRGTRTRGKARTAGARDAWWARAAARAADWARDWDW 781

Qy        521 TGAGAATGTATCAGTCGCTGTTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCT 580
              |   |: ::   |:    |: :|:  : :|:: :   |  ::: |||  ||:|  | : |
Db        780 TATTARRDWTRTARARWRTDAWARTADTDAWRADTTWAAADRRTTTAWRTADTTTTWRWT 721

Qy        581 TTGGACTATATGCTACGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTT 640
              :|: | | | |  ::  | :   |||: ||:  | : :::  ::  |::| |   || :||
Db        720 DTDTARTTTTTWTWWTTTADTWTATAKRTAWRTADTDRDRRRWTTTRDTDTARTTAAKTT 661

Qy        641 CATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAA-AGCATAAAAATAAATATACT 699
                | ||::|  |  : |   :         ||| :: |: |: | :  :: :|:| |: |
Db        660 TTAARAARWARATATWTATTTWAAAAAAAAAATADWRARTARRARRTTWRARARTTTRTT 601

Qy        700 GAAGGTGGTATTTCATGGGTTAATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGT 759
                  :  |:||                  | |||  | |           |  :| |   |
Db        600 TTATAADTTWTTNNNNNANNNNNNTNNNTNTTTNNTNTNTNNNNNNTTTTNTVTTTNTNT 541

Qy        760 TGGATGTCAGGGTATTTTATTTATTTTGTT 789
              |  | |     | | || ||   ||  ||
Db        540 TNTNTNTNNTTNTTTNTTNTTNTNTTNNTT 511


RESULT 13
CN757547
LOCUS       CN757547                 894 bp    mRNA    linear   EST 20-MAY-2004
DEFINITION  ID0AAA1DE11FM1 ApMS Acyrthosiphon pisum cDNA clone ID0AAA1DE11 5',
            mRNA sequence.
ACCESSION   CN757547
VERSION     CN757547.1  GI:47531470
KEYWORDS    EST.
SOURCE      Acyrthosiphon pisum (pea aphid)
  ORGANISM  Acyrthosiphon pisum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
            Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.
REFERENCE   1  (bases 1 to 894)
  AUTHORS   Hunter,W., Martinez-Torres,D., Rahbe,Y., Sabater-Munoz,B.,
            Stern,D., Tagu,D. and Wincker,P.
  TITLE     An expressed sequence tags database for the pea aphid Acyrthosiphon
            pisum
  JOURNAL   Unpublished (2004)
COMMENT     Contact: D. Tagu
            INRA Rennes
            UMR BiO3P, BP 35327, F-35653 Le Rheu Cedex France
            Tel: +33.2.23.48.51.65
            Fax: +33.2.23.48.51.50
            Risk of contamination by bacterial sequences from obligatory
            (Buchnera) or facultative endosymbionts. These sequences were
            obtained in the frame of the International Consortium of Aphid
            Genomics in collaboration with Genoscope
            PCR PRimers
            FORWARD: CAGGAAACAGCTATGACC
            Plate: 1  row: E  column: 11.
FEATURES             Location/Qualifiers
     source          1. .894
                     /organism="Acyrthosiphon pisum"
                     /mol_type="mRNA"
                     /cultivar="developmentstage"
                     /db_xref="taxon:7029"
                     /clone="ID0AAA1DE11"
                     /tissue_type="whole insect"
                     /dev_stage="nymphs and adults (parthenogenetic females)"
                     /lab_host="XL1-Blue"
                     /clone_lib="ApMS"
                     /note="Vector: pBS-SK minus; Site_1: EcoRI; Site_2: XhoI;
                     Sample name: ID0AAA ; Plant growth place: Department of
                     Ecology & Evolutionary Biology, Princeton University ;
                     Soil conditions: Soil ; Sowing date: 01/06/1999 ;
                     Harvesting date: 01/06/1999 ; Stress date: no stress ;
                     Description: Aphids inoculated on one-week old Vicia faba
                     under non-sterile conditions.  All parthenogenetic stages
                     and both winged and wingless adults were collected for
                     library construction. ; experimental condition: long
                     photoperiod (16-hr light/8-hr dark at 18 c)"
ORIGIN

  Query Match             3.8%;  Score 51.4;  DB 7;  Length 894;
  Best Local Similarity   49.8%;  Pred. No. 1.1;
  Matches  130;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy         45 TTATGTTATTTTTGGTAAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAAT 104
              || | || | |||  |||  || |       |||||  |   || | ||| |||   | ||
Db          4 TTTTTTTTTCTTTTTTAAAGTATTTGGAAACTTAACTGAACTATTTTTATTAAGACTTAT 63

Qy        105 ATCACACCATTTATCTAAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGC 164
              | ||  | ||   | || ||||  |||    ||||||| |  | |  |    |      |
Db         64 AACAATCTATAACTTTATAGACTTACAACATTATTATTAGATTTGTTTTGTTATTAACAC 123

Qy        165 ACTACAGATAATATTTGAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGA 224
              |  |   ||| ||  |   || ||  | | |    |||| | |   ||          |
Db        124 ATAATTAATATTAAATTTGTTTGGTTTTTAATGTATAATTAATACTTTATAATATATTAT 183

Qy        225 AATGTCAGCGTTAAAAATATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAA 284
              |   |    ||| ||||||| |       |  ||| ||||   | ||| ||   ||
Db        184 ATATTTTATATTATAATATGAATTGAGGATAAATGATGTTATTTATAAAAATATTGATGG 243

Qy        285 GCAACGTTACCTATCGTTATT 305
              || | ||  ||||||   |||
Db        244 GCTATGTCTCCTATCCCAATT 264


RESULT 14
```

```
CNS000B8
LOCUS       CNS000B8                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR01A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063632
VERSION     AL063632.1  GI:4938680
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR01A24"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             3.8%;  Score 51;  DB 10;  Length 1101;
  Best Local Similarity   34.3%;  Pred. No. 1.3;
  Matches  194;  Conservative   78;  Mismatches  291;  Indels    2;  Gaps    1;

Qy        775 TTTATTTATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGG 834
              | ||| |::  |   :|  |||| |||: :|     :: ||    |   |::  | |: |
Db        493 TATATMTMHMGTHVCMTTTCCCCWATTRAWTRATTTDRATTGCTTGARTTRYTGGGARAG 552

Qy        835 CAGTTAGGGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATA 894
                || ::||    |     |||    |||:|::  :     :   :  :: ||| ::|
Db        553 TGWTTKRKGAAAATTRATTTATTGGAAAATWTDDRSMATTTRRTTTRRGHSYCTTTRRTG 612

Qy        895 TCCACTAAATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGT 954
                  :| |    :   ||   ::|  || ||:     |    | |  |  : || |
Db        613 ATTTGVRATTTGTGRTTKTGWTTRRTTGTGTTTDGTKTGATTTTGTGTTGGTKRTTGTGA 672

Qy        955 AAATCGTTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGT 1014
              : | |   |||  |    |  ||  :| :|  || :||   ||:  |  ||: | |  |
Db        673 TWTTGGGRAAAATATTTTTTTKTGKTWTTDTTTTTGWTTWATTTKGDA--ATWTAATTTT 730

Qy       1015 GTATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTG 1074
               |:|: ||::  :|:|  : :|| :: |||| | :    ::  || :  |    |: ||
Db        731 TTWTWTATWWWTWTWTWAWAWTTTWWTTTATTTTRAATTWRATTTTKGTAATTTTWATTT 790

Qy       1075 GGATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGC 1134
               : ||      ||:||  :||:|:| |:::   ::| ||      |    || |   ||
Db        791 ADTTTTTTWATTTWTTATWTTWTWTATWWWTTTWWTTTTTWATTTATTTTTTTTWATTTT 850

Qy       1135 TTTGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATG 1194
              |||   |   | | || |    ||  |:     |||:    :  |    :  : |  |||
Db        851 TTTTTTAWTAAAAAAAAATTAAATTTAWWAWTTAAAWTTTTWATTTTTWATWWTTAAATT 910

Qy       1195 GCTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTAC 1254
                : || | ::| |  |   | || :|| ||   |  ||    |  |       : |:
Db        911 ATATWTTTWATWWTTAWAAWTTTATTTWTTTTTTTTTTATTTTTTTTTTTTTWTAWATWT 970

Qy       1255 ATGTTAATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTT 1314
              || || |  | |     |   |     |||:: ||| |   | |:| || |||  | |||
Db        971 ATWTTTAAWTTTATTAAAWATATTATWTTAWWATTTTTATATTAWAATTTTATTTTTTT 1030

Qy       1315 AAAAGATTCAAGAGTTCTTATGAGT 1339
              |  |  :: :| : |: ||||  | |
Db       1031 ATTAWTWWAWATWATWTTTATAATT 1055


RESULT 15
CNS010MP/c
LOCUS       CNS010MP                734 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN04L20 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL099163
VERSION     AL099163.1  GI:5610774
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN04L20"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             3.8%;  Score 50.8;  DB 10;  Length 734;
  Best Local Similarity   37.0%;  Pred. No. 1.4;
  Matches  153;  Conservative   63;  Mismatches  195;  Indels    2;  Gaps    2;

Qy        636 ATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAAGCATAAA-AATAAAT 694
              ||||: ||:  |  |||| |||||||| :   || | | || |||  ||||| ||: |||
Db        481 ATTTKTATRNTATAATTTAAAAAAAATGDGTGTGAAGTATTAAAAATATAAAGAAWTAAT 422

Qy        695 ATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCTATGCAATGGCGAATTGCTC 754
              | :   |     |: ||| |||   | || |::|||::  : || ||| :    ||
Db        421 AAWAATATTTGTGWTTTTTATGTTGTTATAAWWTATKWATWTTGTAATAKTWTTTTATGT 362

Qy        755 TAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTATGACCCCCATTGCATTCAAATATT 814
              :   ||  :| | |     ::||:|||||:|||| |::|         ||| |||    |:|
Db        361 KTTWTTTTWTATTTAWTWWTTWTATTTWTTTTTTWWTTKATAWTTTTGTATTWTTTTDTG 302

Qy        815 TCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTTTAACATTATGCAATATGGTAATGG 874
              :      ||    :|      | ||   | ||: ||| :  :|||  :::::: |  |:
```

```
Db        301 KWWAAATTAATATWGWTTTTKATTTTTTGTTGWWTTTTW-TKTATAWWWWWWKTTTTTKW 243

Qy        875 CTACGGGCCTGGCTTGGATATCCACTAAATATCCAAAATGGGGAGTAATGGTTTCCAACA 934
              :  |   ||  |:    |  | :|:|    : |    :::  :  : :  ||
Db        242 WATATTKTWTTTKTTAAAWTAATATAATWTWTTWTTWTTTTWAWKWTTWATWAWAAAAAT 183

Qy        935 AACAGCTTGCGGAACTGAGTAAATCGTTCAAAAGTGCAGTAATGCAATCATCCTTTTTTG 994
              :  |: ::    ||   | |   |   | ||||    | :  :: ||  |    |:  |
Db        182 WTRAKKWWWTTTAAAAAAATTTTTAAATKAAAAAAATATWTTKDAAAAKAAAAATWAATA 123

Qy        995 TCTTGACAGGATTAACTGGTGTATACATTTCATTATGGTTATTGAAATTATCT 1047
              | :|::    :|| : |  | | :| |||| | : |: || :|  :|||:| |
Db        122 TTWTKDWTAWWTTTWWTTTTTTKKAAATTTGAAWKTKTTTTWTATWATTWTAT 70
```

Search completed: May 12, 2006, 03:30:05
Job time : 4188.12 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 19:15:19 ; Search time 183.407 Seconds
(without alignments)
12996.800 Million cell updates/sec

Title: US-09-423-093B-3_COPY_1019_2359
Perfect score: 1341
Sequence: 1 atgaatcgtattattagaat.....tcaagagttcttatgagtaa 1341

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	56	4.2	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	2	54	4.0	1141	3	US-09-806-708B-22	Sequence 22, Appl
	3	49	3.7	32392	3	US-09-662-254B-27	Sequence 27, Appl
c	4	45.4	3.4	1071	3	US-09-710-279-1753	Sequence 1753, Ap
c	5	45.4	3.4	1164	3	US-09-134-001C-369	Sequence 369, App
c	6	45.4	3.4	3053	3	US-09-710-279-3389	Sequence 3389, Ap
	7	45.4	3.4	3159	3	US-09-710-279-4027	Sequence 4027, Ap
c	8	44.6	3.3	2110	3	US-09-419-459-1	Sequence 1, Appli
c	9	44	3.3	832	3	US-09-621-976-2813	Sequence 2813, Ap
	10	44	3.3	1664976	3	US-08-916-421B-1	Sequence 1, Appli
	11	44	3.3	1664976	3	US-09-692-570-1	Sequence 1, Appli
	12	43	3.2	71251	3	US-09-949-016-15332	Sequence 15332, A
c	13	42.4	3.2	1362	3	US-09-248-796A-7845	Sequence 7845, Ap
c	14	42.2	3.1	601	3	US-09-949-016-135624	Sequence 135624,
	15	42.2	3.1	187595	3	US-09-949-016-15546	Sequence 15546, A
c	16	41.2	3.1	32392	3	US-09-662-254B-27	Sequence 27, Appl
c	17	41.2	3.1	254366	3	US-09-822-871-3	Sequence 3, Appli
	18	40.8	3.0	2163	3	US-09-662-254B-38	Sequence 38, Appl
c	19	40.8	3.0	50000	3	US-09-662-254B-24	Sequence 24, Appl
	20	40.8	3.0	50383	3	US-09-949-016-17600	Sequence 17600, A
c	21	40	3.0	601	3	US-09-949-016-18539	Sequence 18539, A
c	22	40	3.0	601	3	US-09-949-016-166322	Sequence 166322,
	23	40	3.0	29769	3	US-09-949-016-16422	Sequence 16422, A
c	24	40	3.0	30820	3	US-09-949-016-17145	Sequence 17145, A
	25	40	3.0	157644	3	US-09-949-016-16179	Sequence 16179, A
	26	40	3.0	157644	3	US-09-949-016-16180	Sequence 16180, A
c	27	39.8	3.0	41182	3	US-09-949-016-17457	Sequence 17457, A
c	28	39.6	3.0	601	3	US-09-949-016-18540	Sequence 18540, A
c	29	39.6	3.0	601	3	US-09-949-016-166323	Sequence 166323,
	30	39.6	3.0	1428	3	US-09-543-681A-2673	Sequence 2673, Ap
	31	39.6	3.0	1689	2	US-07-991-867B-41	Sequence 41, Appl
	32	39.6	3.0	1689	2	US-08-544-332-41	Sequence 41, Appl
	33	39.6	3.0	1689	3	US-09-370-861A-41	Sequence 41, Appl
	34	39.6	3.0	1947	3	US-09-370-861A-74	Sequence 74, Appl
c	35	39.6	3.0	1947	3	US-09-662-254B-55	Sequence 55, Appl
	36	39.6	3.0	8457	2	US-07-991-867B-1	Sequence 1, Appli
	37	39.6	3.0	8457	2	US-08-544-332-1	Sequence 1, Appli
	38	39.6	3.0	8457	3	US-09-370-861A-1	Sequence 1, Appli
c	39	39.6	3.0	32137	3	US-09-949-016-16548	Sequence 16548, A
	40	39.6	3.0	50000	3	US-09-662-254B-26	Sequence 26, Appl
	41	39.6	3.0	640681	3	US-09-790-988-1	Sequence 1, Appli
c	42	39.4	2.9	286	3	US-09-328-111-30	Sequence 30, Appl
	43	39.4	2.9	1301	3	US-09-774-528-160	Sequence 160, App
	44	39.4	2.9	1301	3	US-10-120-988-160	Sequence 160, App
	45	39.2	2.9	832	3	US-09-621-976-2813	Sequence 2813, Ap

ALIGNMENTS

RESULT 1
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 4.2%; Score 56; DB 3; Length 1141;
Best Local Similarity 11.1%; Pred. No. 0.0019;
Matches 122; Conservative 401; Mismatches 556; Indels 17; Gaps 3;

```
Qy          79 ACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCTAAAGACGCACAGGGCTAT 138
               :::::::|:   :    :::  : :::: ::   ::: |  |  : : : : : : :  :
Db          53 MSKSRKWTWARMYCKYRRWYNNKSRWWKGWYKKKWYBCANNTSBRYHARRWKDMKTAYBM 112

Qy         139 TATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTTGAATTGGGGCTATCAACG 198
               |:|  :    : :::: :: ::: : :::: |   |:|     : :::: : :::
Db         113 TMTNKWGKTGWRHRYWRWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNA 172

Qy         199 GTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTATTCTGAACGA 258
                 :::  :::::   ||: :  : :   :::  : :   : ::: : :: |:: :: ::
Db         173 TGWDDDTKYHMWNNNGCBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDD 232

Qy         259 GATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCGGTTGGCAATA 318
               ::: : :::|: :::  ::|:: |  :::   : : : :: :          : :  |
Db         233 MCKRKVRRWVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRA 292

Qy         319 AAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGTATGTT 378
               :   |::      : : ||:: |  :|      :  :::: :  :| :|: :
Db         293 KYWGWNRABVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNN 352
```

```
Qy        379 TTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGGTTATTATTAACA 438
Db        353 NNWTYKKARHBARWDWVWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWA 412

Qy        439 ATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGG 498
Db        413 WYWKMDMDWBGTYNNNNNGGRTYYGWTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTW 472

Qy        499 TTAATTACTGAT-----GTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATAT 553
Db        473 WKMKTYWNNCYWKSMTNGKSHRBAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBV 532

Qy        554 TGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTT 613
Db        533 CSKWWNNYAAWYTKSSWNYTSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTR 592

Qy        614 CAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAAT 673
Db        593 WWBWSHTWBHBRAGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNNN 652

Qy        674 CTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTC 733
Db        653 NNATCARDDYYAASRWYAMANAKWYY-------YKBAANNAYYTHANNWWGCWNNATDTR 705

Qy        734 CTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTATGA 793
Db        706 RTMWKNNNNNNAGTWKNNNNNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADAA 765

Qy        794 CCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTTTAA 853
Db        766 ABTTDKRNNGAYTKYTTTNNNNTYRGVVTNTAARDGWANNNNNNNNNNNNNNGWSDMWVT 825

Qy        854 CATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACTAAATATCCAAAAT 913
Db        826 WWAYANYGTNNNNNNNNNNAYAWWTNKWYYTTDDRWRBAYTNNNNNNRMAYYGAYADDYA 885

Qy        914 GGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCGTTCAAAAGTGCAG 973
Db        886 YYMSDTCDAWMKWDATKMNNA-----TTYNRGTAWRTNNNNNNMTMKTKYYBHAAWNNNN 940

Qy        974 TAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATACATTTCATTATGGT 1033
Db        941 NNGKMCTAHTWWVCKATKTTKGCWMNCTTTCRKYKNNCTWYTWMTTTRTTWYAATRWKTN 1000

Qy       1034 TATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTGCAGGATTTTTTCT 1093
Db       1001 NATGSMTRCNATGWKNNNYWTGWKTRWTAYRMATRWMKAWWKVMATGSWNTNSYARWAYK 1060

Qy       1094 TTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCAACCTATATAAGAG 1153
Db       1061 TRAYKGWYYNACAWRWRWGKATCYMTDNAWWTACATSWMATHKYNWHMCKCNNNNNNNNT 1120

Qy       1154 CGCATAAAACTGAAAA 1169
Db       1121 MMRAMAMAAANCDGAR 1136
```

RESULT 2
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 4.0%; Score 54; DB 3; Length 1141;
Best Local Similarity 10.9%; Pred. No. 0.0053;
Matches 101; Conservative 336; Mismatches 479; Indels 7; Gaps 2;

```
Qy        327 TGCAGTAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGTATGTTTTTTTTAC 386
Db       1049 WSCATKBMWWTMKWYATKYRTAWYAMWCAWRNNNMWCATNGYAKSCATNNAMWYATTRWA 990

Qy        387 GCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGGTTATTATTAACAATAGTTAC 446
Db        989 AYAAAKWARWAGNNMRMYGAAAGNKWGCMAAMATMGBWWADTAGKMCNNNNNNWTTDVRR 930

Qy        447 AGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGGTTAATTAC 506
Db        929 MAMKAKNNNNNNAYWTACYNRAATNNKMATHWMKWTHGAHSKRRTRHHTRTCRRTKYNNN 870

Qy        507 TGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATATTGGCAGTAAGCTT 566
Db        869 NNNARTVYWYHHAARRWMNAWWTRTNNNNNNNNNNACRNTRTWWABWKHSWCNNNNNNNN 810

Qy        567 ACTTA------TTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTTCAGGGAC 620
Db        809 NNNNNNTWCHYTTANABBCYRANNNNAAARMARTCNNYMHAAVTTTHTDWCYKTWMNTWY 750

Qy        621 TATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAA 680
Db        749 WDMMTTMBTTTTRNMTTSTNMTNNNNNNMWACTNNNNNNMWKAYYAHATNNWGCWWNNTD 690

Qy        681 GCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCTATGCA 740
Db        689 ARRTNNTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNNNNNNNNNNNNNNSCCTCTR-M 631

Qy        741 ATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTATGACCCCCAT 800
Db        630 MTMRWTMKGDGMTVRKKVKWRDTTCTYVDVWADSWVWWYANWMRCRDVTYTRNNTYCKSY 571

Qy        801 TGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTTTAACATTATG 860
Db        570 AHSYWYWSNNAMWYRRYSARNWSSMARWTTRNNWWMSGBVRMRWAGTMWWRHWNNNNTDT 511

Qy        861 CAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACTAAATATCCAAAATGGGGAGT 920
Db        510 RYYWWWKRWARBTTTVYDSMCNAKSMWRGNNWRAMKMWWAANNDAGAMDHWTYWMGNNTM 451

Qy        921 AATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCGTTCAAAAGTGCAGTAATGCA 980
Db        450 WMRRAWKMMNMAWCRRAYCCNNNNNRACVWHKHKMWRWTWKYMWKAACNNNNBKAMYMRV 391

Qy        981 ATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATACATTTCATTATGGTTATTGAA 1040
Db        390 AWMMYSRDTTNTDWMMWTSDWBWHWYTVDYTMMRAWNNNNNNNWRBCKTTSWMWWMDHMN 331

Qy       1041 ATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTGCAGGATTTTTTCTTTTTATC 1100
Db        330 THCTYGNNTWGSAYBMAAMSMWAAGASNBVTYNWCWRMTYMGKTMTNNNNNNKAWYYRTK 271

Qy       1101 TTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCAACCTATATAAGAGCGCATAA 1160
Db        270 TVAWCNNRYYYDTAVWTBKRNYKYCYAYBWYYBMYMGKHHWBWWRRABHRSWNMWWVKCR 211

Qy       1161 AACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTCTTGACTATAACTACAATGTT 1220
Db        210 NKYMVSWHYHAMRYBKWABAVGCNNNWKDRMAHHHWCATNNNMMWWWYAYMHHMHKKGKA 151
```

```
Qy       1221 GTTTGTTGCATATTTAGAGTACT 1243
              :|  :| :::::: ::  ::::
Db        150 AWTNNKTABRDDHBAHVKTYWYW 128
```

RESULT 3
US-09-662-254B-27
; Sequence 27, Application US/09662254B
; Patent No. 6933145
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bawden, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; TITLE OF INVENTION: Vertebrate Cells
; FILE REFERENCE: UF-221C1XC1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 32392
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-27

Query Match 3.7%; Score 49; DB 3; Length 32392;
Best Local Similarity 52.2%; Pred. No. 0.13;
Matches 109; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```
Qy       1132 TGCTTTGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATA 1191
              | ||||  ||  |||| || | |  || ||  |  |||||| | ||||   | || || |
Db      22647 TTCTTTAAAAGATATAGAAAATCTTATTAATTTAGAAAAATTAGATTGTTCTAATATAAA 22706

Qy       1192 ATGGCTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTC 1251
              |     |||| | |||||    ||   ||||  ||    | |||||  |  |  ||  |
Db      22707 ACTAAATTCTTAAATATAAAACAAAATATGTTAATTTATAAATTAGATTGTTATAGTATA 22766

Qy       1252 TACATGTTAATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATC 1311
              |||| ||| | | |  |  |     ||  ||  | | |  |     || | | ||| ||
Db      22767 AACATTTTATTTTTTATATTAAATTTGTAATTGAAATATTATAACATAAATAAAATATTC 22826

Qy       1312 TTTAAAAGATTCAAGAGTTCTTATGAGTA 1340
              | | |  ||  | |  |  |||  ||
Db      22827 TATTATGTATAATATATCTACTATTCTTA 22855
```

RESULT 4
US-09-710-279-1753/c
; Sequence 1753, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1753
; LENGTH: 1071
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-1753

Query Match 3.4%; Score 45.4; DB 3; Length 1071;
Best Local Similarity 51.8%; Pred. No. 0.47;
Matches 103; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

```
Qy       1017 ATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGG 1076
              |||  ||||||  |||| | | |    |||||    |||  |||         | |||  |
Db       1026 ATAGTTTTCATATTGGTCAGTAATTAAATCTGAACTAAATCTTGCTTCAATATCTTTCAG 967

Qy       1077 ATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTT 1136
                |    ||||   | |||| || |    |     ||| |||     |  |||| |  ||
Db        966 CATTTGTGATTGCATTTTTTGATATAACTCTGGATTTGATAAAAGTTTAATAGCATATTT 907

Qy       1137 TGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGC 1196
              |||| | | | ||  | | | || | ||   | ||   || |  | ||||| | ||| |
Db        906 TGCAGCTTGTGTACTATCCCCTATATCTACTATAAATCCAGTTTCTTCATGTCTGATGAC 847

Qy       1197 TCTCTTGACTATAACTACA 1215
                || |    | || ||
Db        846 CTCTTTAATACCACCTGCA 828
```

RESULT 5
US-09-134-001C-369/c
; Sequence 369, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 369
; LENGTH: 1164
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-369

Query Match 3.4%; Score 45.4; DB 3; Length 1164;
Best Local Similarity 51.8%; Pred. No. 0.48;
Matches 103; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

```
Qy       1017 ATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGG 1076
              |||  ||||||  |||| | | |    |||||    |||  |||         | |||  |
Db       1119 ATAGTTTTCATATTGGTCAGTAATTAAATCTGAACTAAATCTTGCTTCAATATCTTTCAG 1060

Qy       1077 ATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTT 1136
                |    ||||   | |||| || |    |     ||| |||     |  |||| |  ||
Db       1059 CATTTGTGATTGCATTTTTTGATATAACTCTGGATTTGATAAAAGTTTAATAGCATATTT 1000

Qy       1137 TGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGC 1196
              |||| | | | ||  | | | || | ||   | ||   || |  | ||||| | ||| |
Db        999 TGCAGCTTGTGTACTATCCCCTATATCTACTATAAATCCAGTTTCTTCATGTCTGATGAC 940

Qy       1197 TCTCTTGACTATAACTACA 1215
                || |    | || ||
Db        939 CTCTTTAATACCACCTGCA 921
```

RESULT 6
US-09-710-279-3389/c
; Sequence 3389, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US

```
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3389
;   LENGTH: 3053
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3389

  Query Match             3.4%;  Score 45.4;  DB 3;  Length 3053;
  Best Local Similarity   51.8%;  Pred. No. 0.57;
  Matches  103;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy       1017 ATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGG 1076
              |||  ||||||  |||| | | |    |||||    |||  |||        | |||  |
Db       2945 ATAGTTTTCATATTGGTCAGTAATTAAATCTGAACTAAATCTTGCTTCAATATCTTTCAG 2886

Qy       1077 ATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTT 1136
                |    ||||   | |||| || |    |     ||| |||     |  |||| |  ||
Db       2885 CATTTGTGATTGCATTTTTTGATATAACTCTGGATTTGATAAAAGTTTAATAGCATATTT 2826

Qy       1137 TGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGC 1196
              |||| | | | ||  | | | || | ||   | ||   || |  | |||||| | ||| |
Db       2825 TGCAGCTTGTGTACTATCCCCTATATCTACTATAAATCCAGTTTCTTCATGTCTGATGAC 2766

Qy       1197 TCTCTTGACTATAACTACA 1215
                 || |    | || ||
Db       2765 CTCTTTAATACCACCTGCA 2747


RESULT 7
US-09-710-279-4027
; Sequence 4027, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4027
;   LENGTH: 3159
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-4027

  Query Match             3.4%;  Score 45.4;  DB 3;  Length 3159;
  Best Local Similarity   51.8%;  Pred. No. 0.58;
  Matches  103;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy       1017 ATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGG 1076
              |||  ||||||  |||| | | |    |||||    |||  |||        | |||  |
Db       2357 ATAGTTTTCATATTGGTCAGTAATTAAATCTGAACTAAATCTTGCTTCAATATCTTTCAG 2416

Qy       1077 ATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTT 1136
                |    ||||   | |||| || |    |     ||| |||     |  |||| |  ||
Db       2417 CATTTGTGATTGCATTTTTTGATATAACTCTGGATTTGATAAAAGTTTAATAGCATATTT 2476

Qy       1137 TGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGC 1196
              |||| | | | ||  | | | || | ||   | ||   || |  | |||||| | ||| |
Db       2477 TGCAGCTTGTGTACTATCCCCTATATCTACTATAAATCCAGTTTCTTCATGTCTGATGAC 2536

Qy       1197 TCTCTTGACTATAACTACA 1215
                 || |    | || ||
Db       2537 CTCTTTAATACCACCTGCA 2555


RESULT 8
US-09-419-459-1/c
; Sequence 1, Application US/09419459
; Patent No. 6222028
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jin-Hao
;  APPLICANT: Cheng, Kuo-Joan
;  APPLICANT: Tsai, Cheng-Fang
;  APPLICANT: Chang, Chia-Chieh
;  TITLE OF INVENTION: CELLULASE ENZYMES
;  FILE REFERENCE: 08919/037001
;  CURRENT APPLICATION NUMBER: US/09/419,459
;  CURRENT FILING DATE: 1999-10-15
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2110
;   TYPE: DNA
;   ORGANISM: Piromyces rhizinflata
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3)...(1749)
US-09-419-459-1

  Query Match             3.3%;  Score 44.6;  DB 3;  Length 2110;
  Best Local Similarity   48.9%;  Pred. No. 0.81;
  Matches  149;  Conservative    0;  Mismatches  154;  Indels    2;  Gaps    1;

Qy       1005 ATTAACTGGTGTATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGA 1064
              ||| | |  | |||  |||| ||| |  || |     |  || |  | || |  ||     |
Db       2096 ATTTATTTATTTATTTATTTTATTTTTATTTTATTTAAAATTTTTTACACAATTTTTATA 2037

Qy       1065 GCGGTTTTTGGGATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACAT 1124
                | || |  |||      | |       ||  |||||||  |||| ||| |  | |
Db       2036 ATCTTATTAGCTATTATTTAAAATACATATTCCATCTTTAAAAATTCTTGATTTTAATTA 1977

Qy       1125 TGTAGCTTGCTTTGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATC 1184
                |   ||  ||||       || ||||    ||||     |   ||  ||||  |||
Db       1976 AATTCTTTTTTTTGCTTTAAGTAAAAGAATAAATAATTAAAATTTTATTTCATT--CATT 1919

Qy       1185 ATGTATAATGGCTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTC 1244
              ||   || |   | | ||  || |   || ||| || ||| ||  | ||     || | |
Db       1918 ATTATTATTTTTTATATTTTCTTTTCTTATAATTTTTTTTATTAAAAATAATAAGAAATA 1859

Qy       1245 GAGGTTCTACATGTTAATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTA 1304
                |  || || |||  | | |||      |      ||   || |||||  |  |  |  ||
Db       1858 AAAATTTTAAATGACATTTTATTAGAAATATGTGTATATAAATTTTTCTATTTTATTTTT 1799

Qy       1305 TATAA 1309
              | | |
Db       1798 TTTTA 1794


RESULT 9
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
```

```
;  SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

  Query Match             3.3%;  Score 44;  DB 3;  Length 832;
  Best Local Similarity   14.7%;  Pred. No. 0.94;
  Matches   46;  Conservative  150;  Mismatches  110;  Indels    6;  Gaps    2;

Qy        429 ATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGA 488
              || | |||  ||| ||    ::|::  ::::||: :|: ::::::|  :::: :: :  |
Db        381 ATAAATAATTATATTTTGTYYWTWWKTWWYWTTYTWTRMWWWMKKKARRWYYWWKSTYACA 322

Qy        489 AGGGAGTGGGTTAATT--ACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCT 546
               : :  | | :: :::  ::|::: :  ::  :  |:: |: |:   :  ::   ::::
Db        321 SRYRKYTWGWWWYMWKRMMSTRWYCYMCWKCCMYRGRRCAWYTMARGRMWSYAWGKWKSM 262

Qy        547 GGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATA 606
              :: : : |: : ::   ::|:  |  |: :   ::::::   | ::  : |::|::  |:
Db        261 RSAMSMCTRMYYKKGSTYWTMKCTCATWCYWYWKYWKRMWSKTCWSGSRGGYMTSYTSTR 202

Qy        607 GCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTT- 665
              :: ::: ::   :|::: ::: |::::|: ::::    |::: : :::||: :::: ::
Db        201 SYSMYWASWMYTMCWWWGRWWSTYWYMAWGKKWWRYATTWRRAMMWWWAAWTMMWYMWWA 142

Qy        666 ---CCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAA 722
                 :: | |::  | :::|::::::    ::|: |   |:: :::: : : :: ::  ::
Db        141 WCMSSRGAAMYRRTMMMWGYRYWWRKKSYRRTRCAWAYAWKTKRSYYWCWRWKWKRCMMM 82

Qy        723 TGAAATATTTCC 734
                :|:   :|::
Db         81 MMMAMAYGKTMM 70


RESULT 10
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
;  APPLICANT: Bult et al.
;  TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
;  TITLE OF INVENTION:  jannaschii
;  FILE REFERENCE: PB275
;  CURRENT APPLICATION NUMBER: US/08/916,421B
;  CURRENT FILING DATE:  1997-08-22
;  PRIOR APPLICATION NUMBER: US 60/024,428
;  PRIOR FILING DATE: 1996-08-22
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 3.3%; Score 44; DB 3; Length 1664976;

Best Local Similarity 46.9%; Pred. No. 3.8;
Matches 137; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

```
Qy        92 TAATAATGTTAATATCACACCATTTATCTAAAGACGCACAGGGCTATTATTATACATTTA 151
             ||||| | |||| | | |  ||| | | | | | | | |||| || || | |
Db    651652 TAATATTTTTAAACTATCTTTGTTTTTAGATTTAGTGATTGTATTATTCAAATCCAAATT 651711

Qy       152 ATTCAGTAGTGGCACTACAGATAATATTTGAATTGGGGCTATCAACGGTAATCATTCAAT 211
             ||| ||   | |||     |  |||| ||||| || | |     | | |
Db    651712 ATTTAGAGAAGAAACTGTCTTAACAGTTTGTATTGGATTTAGCGAAATATTTGTTGATTT 651771

Qy       212 TCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTATTCTGAACGAGATATTATAGGTG 271
                 |   ||  |  || |  || |||| |  | | |  ||  |||| ||  ||
Db    651772 ATTTTCATTATTTACAGTTATTATTGAAATTTTTGAAGTATTTACTTGATATGATGAGTT 651831

Qy       272 AAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCGGTTGGCAATAAAATGGTATGCAG 331
             |  |  || || |  |  || |||  |||  | | |  |  ||| ||  |  |||
Db    651832 ATTGACAATATCACTATTATTTCCTCCATTTAAATTGTTATGACTTAACATTACAAACAG 651891

Qy       332 TAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGTATGTTTTTTT 383
             |  |  |  | ||| ||||| |  |  || ||  ||| ||| ||
Db    651892 TCCAAATATAAGAATTAAAATTAAAATAAACGCCAATACCATATTTTTAATT 651943
```

RESULT 11
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6797466
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; CURRENT FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)

; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:

```
;   NAME/KEY: misc_feature
;   LOCATION: (1313224)..(1313224)
;   OTHER INFORMATION: n equals a, t, c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1349473)..(1349473)
;   OTHER INFORMATION: n equals a, t, c, or g

  Query Match             3.3%;  Score 44;  DB 3;  Length 1664976;
  Best Local Similarity   46.9%;  Pred. No. 3.8;
  Matches  137;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy         92 TAATAATGTTAATATCACACCATTTATCTAAAGACGCACAGGGCTATTATTATACATTTA 151
              ||||| | ||||  |  |    ||| |  |   |   |  |   ||||  || ||  |
Db     651652 TAATATTTTTAAACTATCTTTGTTTTTAGATTTAGTGATTGTATTATTCAAATCCAAATT 651711

Qy        152 ATTCAGTAGTGGCACTACAGATAATATTTGAATTGGGGCTATCAACGGTAATCATTCAAT 211
              ||| ||     | |||       |   |||| |||||   || | |      |  |     |
Db     651712 ATTTAGAGAAGAAACTGTCTTAACAGTTTGTATTGGATTTAGCGAAATATTTGTTGATTT 651771

Qy        212 TCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTATTCTGAACGAGATATTATAGGTG 271
              |     ||  |  || |   || |||| |    | | |  ||  ||||| ||  ||
Db     651772 ATTTTCATTATTTACAGTTATTATTGAAATTTTTGAAGTATTTACTTGATATGATGAGTT 651831

Qy        272 AAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCGGTTGGCAATAAAATGGTATGCAG 331
              |  |  || || | |  | || |||   ||| | | |  |  ||| ||   |  |||
Db     651832 ATTGACAATATCACTATTATTTCCTCCATTTAAATTGTTATGACTTAACATTACAAACAG 651891

Qy        332 TAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGTATGTTTTTTT 383
              |   |   |    | ||| ||||| |      || ||    ||| |||  ||
Db     651892 TCCAAATATAAGAATTAAAATTAAAATAAACGCCAATACCATATTTTTAATT 651943


RESULT 12
US-09-949-016-15332
; Sequence 15332, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15332
;   LENGTH: 71251
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15332

  Query Match             3.2%;  Score 43;  DB 3;  Length 71251;
  Best Local Similarity   50.2%;  Pred. No. 3.6;
  Matches  106;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy       1123 ATTGTAGCTTGCTTTGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCA 1182
              ||| ||  |  ||||   ||||  ||||    ||||| |    | |||||  |   |
Db      55391 ATTATACATGACTTTCTTACCTTAATAACCAGGCATATATAAAAGGAAATGCTAACAGTT 55450

Qy       1183 TCATGTATAATGGCTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTAC 1242
              |  | |||  | | |  || | | |  ||||| ||| | || |  |||||||  ||||
Db      55451 TTCTCTTTAAAGTCACATTTCAGTTTTCCTACACTGTAGGTTCTATCATATTTTGAGTCG 55510

Qy       1243 TCGAGGTTCTACATGTTAATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACT 1302
              |  |  ||| |  |   | | | |    | |  |  |  |  | |  | ||| |      |
Db      55511 TGCAACTTCCATTTTACATTTTTTTTTCCCCACAACTTTTGCTTTCCTTTTTCCTCTTTT 55570

Qy       1303 TATATAATCTTTAAAAGATTCAAGAGTTCTT 1333
              | |      |||  ||||      ||| ||
Db      55571 TCTCCTGCATTTTGAAGAAAAGGAGGTTTTT 55601


RESULT 13
US-09-248-796A-7845/c
; Sequence 7845, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 7845
;   LENGTH: 1362
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-09-248-796A-7845

  Query Match             3.2%;  Score 42.4;  DB 3;  Length 1362;
  Best Local Similarity   49.1%;  Pred. No. 2.4;
  Matches  112;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        541 TTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCT 600
              | |||  |||||  |  |  || | | ||  |||  | | |    || | | | | | |
Db        671 TCAGCATGTATAGTATCGATATGATGACCAATTTTTTGATCAATACCAAAAGATCCACCA 612

Qy        601 GCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAA 660
              |  |  |  |   |||   | |  | ||  |  |   | |||  |  || ||||   | |
Db        611 GGGAAGGGAACAGCAGCACCAACAAGATAATGTAATCTATCAGTTGGGTTTTTTTTCACA 552

Qy        661 ATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTT 720
              |  |    |  ||||||    ||  | | | ||| ||| | |||| |  ||     | |
Db        551 ACGTAAAAGTCTTCTTTAGCCAATTTAGAAACATAAACTAATGGTGATTCTTTCAATGGT 492

Qy        721 AATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCA 768
              || |  |||| |  |||     |     ||| |  || |   ||| | |
Db        491 AAAGTCATATCTAATATTTCTTTCGTGATTTCATTACGAATGATTTTA 444


RESULT 14
US-09-949-016-135624/c
; Sequence 135624, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 135624
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-135624
```

```
  Query Match             3.1%;  Score 42.2;  DB 3;  Length 601;
  Best Local Similarity   55.0%;  Pred. No. 2.3;
  Matches   83;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        968 GTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATACATTTCAT 1027
              || ||||| || ||   |  ||  || ||||    |  | || |  | | |||| ||  |
Db        474 GTCCAGTAGTGAAACGGTATTTCATTTTCTTCCTCGTGTAAATTTTTATTTACAATTGCT 415

Qy       1028 TATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTGCAGGATT 1087
               | |||||| || |||     |||     ||||| |    ||||  |    |  |||
Db        414 GCTTGTTATTAAATTTAGTAATTTCGTTTTTTGGCAACTTATTTTCTGTTGAGTGGGAAC 355

Qy       1088 TTTTCTTTTTATCTTTAGCAATTATTGGTAA 1118
              || ||| ||| |||||   ||  |  |  ||
Db        354 TTGTCTGTTTCTCTTTTCTAAAAAAGGTGAA 324


RESULT 15
US-09-949-016-15546
; Sequence 15546, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15546
;   LENGTH: 187595
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(187595)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15546

  Query Match             3.1%;  Score 42.2;  DB 3;  Length 187595;
  Best Local Similarity   55.0%;  Pred. No. 6.5;
  Matches   83;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        968 GTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATACATTTCAT 1027
              || ||||| || ||   |  ||  || ||||    |  | || |  | | |||| ||  |
Db      85917 GTCCAGTAGTGAAACGGTATTTCATTTTCTTCCTCGTGTAAATTTTTATTTACAATTGCT 85976

Qy       1028 TATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTGCAGGATT 1087
               | |||||| || |||     |||     ||||| |    ||||  |    |  |||
Db      85977 GCTTGTTATTAAATTTAGTAATTTCGTTTTTTGGCAACTTATTTTCTGTTGAGTGGGAAC 86036

Qy       1088 TTTTCTTTTTATCTTTAGCAATTATTGGTAA 1118
              || ||| ||| |||||   ||  |  |  ||
Db      86037 TTGTCTGTTTCTCTTTTCTAAAAAAGGTGAA 86067


Search completed: May 12, 2006, 03:51:47
Job time : 192.407 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:31:18 ; Search time 858.623 Seconds
(without alignments)
12915.142 Million cell updates/sec

Title: US-09-423-093B-3_COPY_1019_2359
Perfect score: 1341
Sequence: 1 atgaatcgtattattagaat.....tcaagagttcttatgagtaa 1341

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1341	100.0	12441	3	US-09-423-093A-3	Sequence 3, Appli
2	50.6	3.8	19131	7	US-10-221-714A-442	Sequence 442, App
3	50.4	3.8	8923	7	US-10-221-714A-62	Sequence 62, Appl
4	50.2	3.7	17234	7	US-10-433-793-48	Sequence 48, Appl
5	49.8	3.7	21537	6	US-10-311-455-1972	Sequence 1972, Ap
6	49.4	3.7	7255	6	US-10-311-455-1251	Sequence 1251, Ap
7	49.4	3.7	12138	6	US-10-311-455-1601	Sequence 1601, Ap
8	49.4	3.7	12138	7	US-10-257-166-115	Sequence 115, App
9	49.2	3.7	8123	6	US-10-311-455-2200	Sequence 2200, Ap
10	49	3.7	32392	8	US-10-706-635-27	Sequence 27, Appl
11	47.6	3.5	6577	6	US-10-311-455-1329	Sequence 1329, Ap
12	47.6	3.5	6577	7	US-10-221-613-181	Sequence 181, App
13	47.6	3.5	14924	6	US-10-311-455-197	Sequence 197, App
14	47.6	3.5	14924	6	US-10-240-452-21	Sequence 21, Appl
15	46.6	3.5	6191	6	US-10-311-455-1190	Sequence 1190, Ap
16	46.4	3.5	6392	6	US-10-311-455-657	Sequence 657, App
17	46.4	3.5	6392	6	US-10-240-485-59	Sequence 59, Appl
18	46	3.4	15373	6	US-10-311-455-440	Sequence 440, App
19	46	3.4	3673778	6	US-10-312-841-2	Sequence 2, Appli
20	45.8	3.4	7625	6	US-10-311-455-1404	Sequence 1404, Ap
21	45.8	3.4	18281	6	US-10-311-455-935	Sequence 935, App
22	45.8	3.4	71997	7	US-10-322-281-426	Sequence 426, App
23	45.6	3.4	3720	5	US-10-172-086-74	Sequence 74, Appl
24	45.6	3.4	3720	7	US-10-311-507-38	Sequence 38, Appl
25	45.6	3.4	3720	8	US-10-480-846-74	Sequence 74, Appl
26	45.6	3.4	3720	8	US-10-473-126-146	Sequence 146, App
27	45.6	3.4	8030	6	US-10-311-455-1306	Sequence 1306, Ap
28	45.6	3.4	17869	6	US-10-311-455-77	Sequence 77, Appl
29	45.6	3.4	17869	7	US-10-257-166-1	Sequence 1, Appli
30	45.6	3.4	18218	6	US-10-311-455-1922	Sequence 1922, Ap
31	45.6	3.4	40862	6	US-10-311-455-2045	Sequence 2045, Ap
c 32	45.4	3.4	1140	7	US-10-282-122A-34407	Sequence 34407, A
33	45.4	3.4	12669	7	US-10-221-613-412	Sequence 412, App
c 34	45.2	3.4	552	4	US-09-925-065A-456703	Sequence 456703,
35	45.2	3.4	18855	6	US-10-311-455-584	Sequence 584, App
c 36	45	3.4	1620	3	US-09-950-051-8	Sequence 8, Appli
37	45	3.4	5641	6	US-10-311-455-1369	Sequence 1369, Ap
38	44.8	3.3	6831	6	US-10-311-455-1460	Sequence 1460, Ap
39	44.8	3.3	7040	5	US-10-172-086-13	Sequence 13, Appl
40	44.8	3.3	7040	7	US-10-221-714A-161	Sequence 161, App
41	44.8	3.3	7040	7	US-10-311-507-47	Sequence 47, Appl
42	44.8	3.3	7040	8	US-10-480-846-13	Sequence 13, Appl
43	44.8	3.3	7040	8	US-10-473-126-173	Sequence 173, App
44	44.8	3.3	7040	8	US-10-473-126-319	Sequence 319, App
45	44.8	3.3	18011	6	US-10-311-455-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-423-093A-3
; Sequence 3, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12441
; TYPE: DNA
; ORGANISM: Salmonella enterica
US-09-423-093A-3

Query Match 100.0%; Score 1341; DB 3; Length 12441;
Best Local Similarity 100.0%; Pred. No. 6.4e-285;
Matches 1341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 ATGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGT 1078

Qy         61 AAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 AAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCT 1138

Qy        121 AAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 AAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTT 1198

Qy        181 GAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 GAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAA 1258

Qy        241 TATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 TATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCG 1318
```

```
Qy        301 TTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 TTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTC 1378

Qy        361 GGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 GGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGG 1438

Qy        421 GCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 GCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCT 1498

Qy        481 GTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1499 GTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTG 1558

Qy        541 TTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1559 TTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCT 1618

Qy        601 GCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1619 GCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAA 1678

Qy        661 ATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1679 ATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTT 1738

Qy        721 AATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 AATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATT 1798

Qy        781 TATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1799 TATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTA 1858

Qy        841 GGGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1859 GGGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACT 1918

Qy        901 AAATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1919 AAATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCG 1978

Qy        961 TTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1979 TTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATAC 2038

Qy       1021 ATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2039 ATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTG 2098

Qy       1081 CAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2099 CAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCA 2158

Qy       1141 ACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2159 ACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGCTCTC 2218

Qy       1201 TTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACATGTTA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2219 TTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACATGTTA 2278

Qy       1261 ATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTTAAAAGA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2279 ATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATCTTTAAAAGA 2338

Qy       1321 TTCAAGAGTTCTTATGAGTAA 1341
              |||||||||||||||||||||
Db       2339 TTCAAGAGTTCTTATGAGTAA 2359
```

RESULT 2
US-10-221-714A-442
; Sequence 442, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 442
; LENGTH: 19131
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-442

Query Match 3.8%; Score 50.6; DB 7; Length 19131;
Best Local Similarity 45.9%; Pred. No. 2.8;
Matches 173; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

```
Qy        136 TATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTTGAATTGGGGCTATCA 195
              |||| | | |      ||   | |  ||  | || | |  |||||  | |         | |
Db         72 TATTGTAAAAAGAAAAAGAAAATGATGTTAATAAAAAGTATATTGTATTGTTATAAGTTA 131

Qy        196 ACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTATTCTGAA 255
                  ||  |   |  | | ||   || | |  | |  || || ||   ||| ||   ||
Db        132 TATAAAAATAAAAATATAGTTATAAATAATAATTTATTGTGAAGATTATATTTTTTATAA 191

Qy        256 CGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCGGTTGGCA 315
              |     ||| | |  ||  ||||| |    |  ||||    |  ||   | | ||   |
Db        192 AGTTTATGTATGGTTTAAGTTAAAATTGTAAATAGTTATAGTTTTTTTGATAGATTATTA 251

Qy        316 ATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGTAT 375
              |  |       |   |   | |   | | ||||  |||| ||  |   |   ||  | |
Db        252 AATATGATAAAATTATGTAAAAAGTAGGTAATGTTATTTATGATTTTTAATATTAGTGAA 311

Qy        376 GTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGGCATGGTTATTATTA 435
              |||| ||||   | ||||||||   ||    |   | | |   |   ||| ||
Db        312 TTTTTATTTAAAAAGAAAGAAGGAAAAGAAAAATAAGTGATACGTATTTTTGGATTGTTT 371

Qy        436 ACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGT 495
              |  ||| ||| |   ||| | ||   | || ||| | ||   | |     |  ||  ||
Db        372 AAGATATTTAAAATATTTTAAATAAATTTTATTTTTTTATAGTATTAGAATATAGAAAGG 431

Qy        496 GGGTTAATTACTGATGT 512
                ||  ||| || | |
Db        432 TAATTTTTTATTGTTTT 448
```

RESULT 3
US-10-221-714A-62
; Sequence 62, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:

```
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 62
;   LENGTH: 8923
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-62

  Query Match             3.8%;  Score 50.4;  DB 7;  Length 8923;
  Best Local Similarity   50.0%;  Pred. No. 2.3;
  Matches  126;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy        621 TATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAA 680
              |||   ||| |   | ||||  | ||  ||||||  ||    |   |  || | | || | |
Db       1228 TATTGGATTATTTTTTTTTTTTGTATATATTTTTAATTTAAAATTTGAATTATTAAGA 1287

Qy        681 GCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCTATGCA 740
              | |  ||| ||| | |    | | |  | | | |||    ||  || |  ||  |    |
Db       1288 GGAATAAATTAATTGTTTAAAGGTTAATTTATAATGAAAAAAATAATTTGTTTATGATAA 1347

Qy        741 ATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTATGACCCCCAT 800
              |    |  |  | | | | | ||   | | ||| || ||| |  | |||  |
Db       1348 TTTTAAAGGTATTTTTAATAGTATTATAAGTTATATTTTTTTTGATATTAAAATTAAATG 1407

Qy        801 TGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTTTAACATTATG 860
              |  |||    | |||  | | ||| |||  |    |  |||  |||  ||    | |||
Db       1408 TAGATTATTTTTTTTATGTGGTATTTATAAAATAAATATAGTAATGAATTATTTAATATT 1467

Qy        861 CAATATGGTAAT 872
              |||||  ||||
Db       1468 AAATATTTTAAT 1479


RESULT 4
US-10-433-793-48
; Sequence 48, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/433,793
;  CURRENT FILING DATE:  2003-06-06
;  NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 48
;   LENGTH: 17234
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-48

  Query Match             3.7%;  Score 50.2;  DB 7;  Length 17234;
  Best Local Similarity   43.6%;  Pred. No. 3.3;
  Matches  328;  Conservative    0;  Mismatches  418;  Indels    7;  Gaps    2;

Qy         41 TTCGTTATGTTATTTTTGGTAAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGT 100
              || ||   |  ||| ||   | | |  ||| |||  ||  |||  |  |  |    || |
Db       2296 TTTGTACGGAAATTATTTAGATGTTTTTATGTGTTGTAGAGGGGATGTTAGTTGAGATTT 2355

Qy        101 TAATATCACACCATTTATCTAAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAG 160
              | | |      || |   ||      |  |  |||| ||  |   |||| |
Db       2356 AGAAAATATTTTTGTTTTGGTAATTGTTAAGGTTTTATTTTTTAAATGTTAAGTTTTAGA 2415

Qy        161 TGGCACTACAGATAATATTTGAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCC 220
              |   | ||    ||  ||| ||  |   |     |     |  ||  |    ||
Db       2416 TTTTAATATTTTTATGATTATAAGGGATTTTGGAGTAGTATTTTTTTTGAGGATTTAAAA 2475

Qy        221 ATGAAAATGTCAGCGTTAAAATATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAA 280
              | |||||| |   |||| | | | ||  |       | || ||||   |   | |||
Db       2476 TTAAAATGTTATTTTTAATAGAGGTTTTGTGGTGGAGATATTTTATTTTTTGTATTAATT 2535

Qy        281 ATAAGCAACGTTACCTATCGTTATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTT 340
              | | |     |||  | |   |  ||           |   |  |||  | | | |  |
Db       2536 AAATGGTTTTTTATTTTTAAGTGATTTTAGAAAGTTTTTTTTTTTTATTGATTTAAATGTG 2595

Qy        341 TGCTAATAATATTAATAGTCGGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCT 400
              |  |  | | |||  | |          ||    | | |||| |      ||| |  |   |
Db       2596 TTTTTTTAAATTTTTGATTTGTTTATTTTATTTTTTTTGTAGTTATAAAGAATAAATT 2655

Qy        401 TAGGTGTACCTTGGCAAGGGGCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTT 460
              |   | |    |    |       | ||||||  | ||||||       |||  | |||
Db       2656 TGTTTTTTTTATTTATAAATAATTTTTTATTAGGAGTAATAGTGGTTATGTTTTTTTTTT 2715

Qy        461 TTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAA 520
              || || |||    | ||| |    ||  || |  |  || | | |  |  ||  | | |
Db       2716 TTTTTTTTTTAAAATTTTTT-----TGTTAGGTTTATTATTTTTTATTTTTTTTAAGATT 2770

Qy        521 TGAGAATGTATCAGTCGCTGTTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCT 580
              |   |||  || |   | || | |  |||| | || | || |   |   ||   | | |
Db       2771 TTTTAATTGATTAAATGGTGGTGGGAGGTAATTCGGTATTATGGAGAAAAATATTTAGTT 2830

Qy        581 TTGGACTATATGCTACGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTT 640
              |   | || ||| || |             |||| ||       | |  || |   | ||||
Db       2831 TGAAATTAAATGTTAGGGGGTTTTGTTTTATTATAGTAGTATTAAGTTTTTTGGTTTTTT 2890

Qy        641 CATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTG 700
              | ||     ||||   |    ||   || |||   || | |    |||    |   |
Db       2891 TAGATTTTGGTTTTTTTATTTATTAATTGAAATGGGTAGATTAGGTGAATTTTGAAGTGG 2950

Qy        701 AAGGT--GGTATTTCATGGGTTAATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAG 758
              || |  ||| ||| |   |||  | | ||||||    |  | |   |  ||| | ||
Db       2951 TAGTTTCGGTTTTTAAAATGTTTTTAAGATATTTTAAGTATATTTTGGTTTTGATATATA 3010

Qy        759 TTGGATGTCAGGGTATTTTATTTATTTTGTTAT 791
              | |  | |   || ||| ||| | |||  || |
Db       3011 TGGTTTTTTTTGGGATTATATGTTTTTATTTGT 3043


RESULT 5
US-10-311-455-1972
; Sequence 1972, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
```

```
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1972
;   LENGTH: 21537
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1972

  Query Match             3.7%;  Score 49.8;  DB 6;  Length 21537;
  Best Local Similarity   46.6%;  Pred. No. 4.5;
  Matches  159;  Conservative    0;  Mismatches  182;  Indels    0;  Gaps    0;

Qy        345 AATAATATTAATAGTCGGTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGG 404
              ||||  || ||| ||   |   |      || |    |||| |     ||      ||
Db        146 AATAGGATAAATTGTTTTTTTTAATAATAATATAAAATTTATGATTTTAGTTAATATATA 205

Qy        405 TGTACCTTGGCAAGGGGCATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCT 464
              | |   || | |    |  |  || || ||   | ||  |   | |||   ||| |
Db        206 TATTTTTTAGAAGATAGATTTTTTTTTTTTGTTATTAAATTGTGTTTTGGGTATGTAAAG 265

Qy        465 TGTTTCTGTACTTTCTGTCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAG 524
              |      ||| ||| |||       | ||  || ||| ||  |||| | | | || |||
Db        266 TTAAAGGGTAGTTTTTGTGATATTATGGTTTGAGTTTTTTTATGATTTAATTTAAGTGAT 325

Qy        525 AATGTATCAGTCGCTGTTAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGG 584
              | |       |  |||||      ||||||  |||| | ||     ||| | |   |
Db        326 ATTTATATTATTATTGTTATTGTTAATATTGTTAGTAGGTTTTTAGTTTATTTGGAATTT 385

Qy        585 ACTATATGCTACGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATA 644
              | | |     |           || | | |   | | ||   |       |||||| | |
Db        386 ATAAAAGTAAAAAATGTATTAAGTTTTGTAGAGTATTAGGGTTAAAAATATATTTTTAAA 445

Qy        645 TAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAAGCATA 685
              | | | |||||| ||||   |  || ||  |||||   |||
Db        446 TGATTGTTTTAATAAAAGCGTTATGAAAAATTTAATTTATA 486


RESULT 6
US-10-311-455-1251
; Sequence 1251, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1251
;   LENGTH: 7255
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1251

  Query Match             3.7%;  Score 49.4;  DB 6;  Length 7255;
  Best Local Similarity   48.1%;  Pred. No. 3.5;
  Matches  140;  Conservative    0;  Mismatches  151;  Indels    0;  Gaps    0;

Qy        501 AATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATATTGGCAGT 560
              ||  |  | || |  |||  |  | || | ||  || | | |    | |||| ||   | |
Db       3382 AAGTTGTTATATTTATATTAATATAAAGAATGGGTTGTTATATTTTTGTATGTTTATATT 3441

Qy        561 AAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTTCAGGGAC 620
                ||  |||| | | | |||  | | |||  |   | || |||   | ||       |
Db       3442 TTATTTTTTTATGATTCGGTTTAAATTTTATTTTTTTTTATGAAATTTTTTTTGATTTTAT 3501

Qy        621 TATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTTAAA 680
              | |  |  | |   | || |  | |    || || | |  |||||  |  ||| ||||
Db       3502 TTTTTTTATATTATTGTTGTTTTTTTTAAATATTTATAGTATTTTATTTAAATTTTTATT 3561

Qy        681 GCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCTATGCA 740
              | |   | ||||| |      | |  |      |   ||| | ||  |  || | |
Db       3562 AATTTATTTTTAATATTTTAGTATTTATTATAGTATGTGGTATGTAGTAGATGATAGGAA 3621

Qy        741 ATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTAT 791
              |||  | |   |  | || |||    |  ||||    ||| ||||||
Db       3622 ATGTTTATTGAATGAATGTAGGAATATATATTATTAATATTAATTTGTTGT 3672


RESULT 7
US-10-311-455-1601
; Sequence 1601, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1601
;   LENGTH: 12138
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1601

  Query Match             3.7%;  Score 49.4;  DB 6;  Length 12138;
  Best Local Similarity   48.7%;  Pred. No. 4.3;
  Matches  163;  Conservative    0;  Mismatches  171;  Indels    1;  Gaps    1;

Qy        957 ATCGTTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGT 1016
              || | |  | || | | |  |  | | ||  |||| | | | ||   |   | |||| |
Db         99 ATTGGTAGAGAGAGTATTTTTATATTTATTTTTTTGTTTGTAGAGTAGGAGGAGTGGTAT 158

Qy       1017 ATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGG 1076
                |  | || |||   | | ||   | |   | |||      |  ||      ||||||  |
Db        159 TTGGAATTGATTTATGGTTTTAGTAATTATTAGTTTTTTAAAGGGTTTTAAGTTTTTAGA 218

Qy       1077 ATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTT 1136
                | |   | | | | | | | | |  | ||  |  |  ||  | | | | ||| ||  |
Db        219 TTAGTTTGTGTGTGTGTATGTTTAGTGTGCGGTGGTATGTTTTTATAATTTAGTTTTGTA 278

Qy       1137 TGCAACCTATATAAGAGCGCAT-AAAACTGAAAAAATGACATTGGCATCATGTATAATGG 1195
                ||  | | |  || ||  |  |  |  |  | |  |||| |  | ||| ||
Db        279 AAGAAGTTTTTTGCGATCGGTTCGTATATAGGTTTTTTAGTTTGGTAGGTTTTATTATTT 338

Qy       1196 CTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACA 1255
                | | || |  |||| |   || || ||   ||  |  ||||  || | | ||  || || |
Db        339 TTATTTTTATAATAATTTTTATATTTTTAAGTGTTTTGTTAGTATATTTGATTTTTTAAA 398
```

```
Qy       1256 TGTTAATGTATGCAGCACTAACGTGGTTATATTTT 1290
              |  |  |||| |    | | |  ||||  ||||||
Db        399 TAGTTTTGTAAGATAAATTGAATTGGTATTATTTT 433


RESULT 8
US-10-257-166-115
; Sequence 115, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
;  TITLE OF INVENTION:  Genes Implicated in Pharmacogenomics
;  FILE REFERENCE: 5013.1011
;  CURRENT APPLICATION NUMBER: US/10/257,166
;  CURRENT FILING DATE:  2002-10-07
;  PRIOR APPLICATION NUMBER: PCT/EP01/07470
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-06-29
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 115
;   LENGTH: 12138
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-115

  Query Match             3.7%;  Score 49.4;  DB 7;  Length 12138;
  Best Local Similarity   48.7%;  Pred. No. 4.3;
  Matches  163;  Conservative    0;  Mismatches  171;  Indels    1;  Gaps    1;

Qy        957 ATCGTTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGT 1016
              || | |  | || | | |  | | | ||  |||| | | | ||   |    | |||| |
Db         99 ATTGGTAGAGAGAGTATTTTTATATTTATTTTTTTGTTTGTAGAGTAGGAGGAGTGGTAT 158

Qy       1017 ATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGG 1076
              |  | || |||   | | ||   | |   | |||     |  ||      ||||||
Db        159 TTGGAATTGATTTATGGTTTAGTAATTATTAGTTTTTAAAGGGTTTTAAGTTTTTAGA 218

Qy       1077 ATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTT 1136
              | |   |  | | | | | | |  | ||  |  |  ||  | | | | ||| ||  |
Db        219 TTAGTTTGTGTGTGTGTATGTTTAGTGTGCGGTGGTATGTTTTTATAATTTAGTTTTGTA 278

Qy       1137 TGCAACCTATATAAGAGCGCAT-AAAACTGAAAAAATGACATTGGCATCATGTATAATGG 1195
                ||  | | |  || ||  |   |  |       | |  |||| |   | ||| ||
Db        279 AAGAAGTTTTTTGCGATCGGTTCGTATATAGGTTTTTTAGTTTGGTAGGTTTTATTATTT 338

Qy       1196 CTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTCTACA 1255
              | | || |  |||| |   || || ||   ||  |  ||||  || | ||  || || |
Db        339 TTATTTTTATAATAATTTTTATATTTTTAAGTGTTTTGTTAGTATATTTGATTTTTTAAA 398

Qy       1256 TGTTAATGTATGCAGCACTAACGTGGTTATATTTT 1290
              |  |  |||| |    | | |  ||||  ||||||
Db        399 TAGTTTTGTAAGATAAATTGAATTGGTATTATTTT 433


RESULT 9
US-10-311-455-2200
; Sequence 2200, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2200
;   LENGTH: 8123
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2200

  Query Match             3.7%;  Score 49.2;  DB 6;  Length 8123;
  Best Local Similarity   46.9%;  Pred. No. 4;
  Matches  153;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy        988 TTTTTTGTCTTGACAGGATTAACTGGTGTATACATTTCATTATGGTTATTGAAATTATCT 1047
              |||||| | |||      || |         ||  ||| ||| |  || ||| || ||| |
Db        144 TTTTTTTTTTTGTTGTATTTTAGGATAACGTAATTTTAATTTTTTTTTTTGTAAATATAT 203

Qy       1048 GGTTCAAACATTGGCGAGCGGTTTTTGGGATTGCAGGATTTTTTCTTTTTATCTTTAGCA 1107
                ||     | |   ||  | |||            |  |||||| | ||||  |||
Db        204 TATTTGTTTAGTAAAGACGGTTTTAATTTTAATTCGTTTTTTTTTTATTTAAATTTTTTT 263

Qy       1108 ATTATTGGTAATCACATTGTAGCTTGCTTTGCAACCTATATAAGAGCGCATAAAACTGAA 1167
                | ||  ||||   || || | ||  ||||     || ||         | |  | | |
Db        264 TATTTTTTAATTGTATAGTTGTTTTGTTTGTTTTTAAATGGAGTATTAAATTAGTTTA 323

Qy       1168 AAAATGACATTGGCATCATGTATAATGGCTCTCTTGACTATAACTACAATGTTGTTTGTT 1227
              ||| | |   |     |      || | | | | | |  | | || |  | ||| ||| ||
Db        324 AAATTAAATGTAATTTAGGTAGTATTTGTTTTTTAGTTTTTTACGATTAAGTTTTTTTTT 383

Qy       1228 GCATATTTAGAGTACTCGAGGTTCTACATGTTAATGTATGCAGCACTAACGTGGTTATAT 1287
              | | || | |       ||||  |  | || |||||  |     |  | |    ||| | |
Db        384 GAAAATATCGGTGTGGTGAGGGTTAAGATTTTAATTTTGTTATTAGTTGTTTGGGTTTGG 443

Qy       1288 TTTGTTCCTCAAACTTATATAATCTT 1313
              |   |  |   | || |   || ||
Db        444 GTAAGTTATTTTAGTTTTTAGATGTT 469


RESULT 10
US-10-706-635-27
; Sequence 27, Application US/10706635
; Publication No. US20050014263A1
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XCZ1
;  CURRENT APPLICATION NUMBER: US/10/706,635
;  CURRENT FILING DATE:  2003-11-12
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 09/662,254
;  PRIOR FILING DATE: 2000-09-14
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
;   LENGTH: 32392
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-10-706-635-27

  Query Match             3.7%;  Score 49;  DB 8;  Length 32392;
```

```
  Best Local Similarity   52.2%;  Pred. No. 8;
  Matches  109;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy       1132 TGCTTTGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATA 1191
              | ||||  ||  |||| || | |  || ||  |  |||||| | ||||   | || || |
Db      22647 TTCTTTAAAAGATATAGAAAATCTTATTAATTTAGAAAAATTAGATTGTTCTAATATAAA 22706

Qy       1192 ATGGCTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAGGTTC 1251
              |      |||| | |||||     ||   |||| ||     | |||||  |  |  || |
Db      22707 ACTAAATTCTTAAATATAAAACAAAATATGTTAATTTATAAATTAGATTGTTATAGTATA 22766

Qy       1252 TACATGTTAATGTATGCAGCACTAACGTGGTTATATTTTGTTCCTCAAACTTATATAATC 1311
              |||| ||| | | |  | |      ||  ||  | | |  |      || | | ||| ||
Db      22767 AACATTTTATTTTTTATATTAAATTTGTAATTGAAATATTATAACATAAATAAAATATTC 22826

Qy       1312 TTTAAAAGATTCAAGAGTTCTTATGAGTA 1340
              | | |   ||   | |  |  |||   ||
Db      22827 TATTATGTATAATATATCTACTATTCTTA 22855


RESULT 11
US-10-311-455-1329
; Sequence 1329, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1329
;   LENGTH: 6577
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1329

  Query Match             3.5%;  Score 47.6;  DB 6;  Length 6577;
  Best Local Similarity   53.8%;  Pred. No. 8.3;
  Matches   98;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy       1054 AACATTGGCGAGCGGTTTTTGGGATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATT 1113
              || |||   |||   |||||| | |   ||  |||||| ||||| | ||||  |||| |
Db       4340 AATATTTTAGAGGTTTTTTTGTGTTAATAGTTTTTTTTTTTTTTTTTTTTAAAAATTTTA 4399

Qy       1114 GGTAATCACATTGTAGCTTGCTTTGCAACCTATATAAGAGCGCATAAAACTGAAAAAATG 1173
              || || |  ||  || || |||| ||     ||  |    |||  | ||       ||
Db       4400 AGTGATTAGTTTAGAGGTTTGTTTGGAATAATGGTATTATTTAATATTATTGGTTTTTTG 4459

Qy       1174 ACATTGGCATCATGTATAATGGCTCTCTTGACTATAACTACAATGTTGTTTGTTGCATAT 1233
              |||  |   | | || ||  | |  |     |||      | | |||| ||| || |
Db       4460 GAATTTGGGTATTTTTTAGTGTGTTTTATATTGTTAATATTTTTTTAGTTTTTTGGATTT 4519

Qy       1234 TT 1235
              ||
Db       4520 TT 4521


RESULT 12
US-10-221-613-181
; Sequence 181, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, ALexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
;  FILE REFERENCE: 5013.1004
;  CURRENT APPLICATION NUMBER: US/10/221,613
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: PCT/EP01/02945
;    DE 10013847.00
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-03-15
;    2000-03-15
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 181
;   LENGTH: 6577
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-181

  Query Match             3.5%;  Score 47.6;  DB 7;  Length 6577;
  Best Local Similarity   53.8%;  Pred. No. 8.3;
  Matches   98;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy       1054 AACATTGGCGAGCGGTTTTTGGGATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATT 1113
              || |||   |||   |||||| | |   ||  |||||| ||||| | ||||  |||| |
Db       4340 AATATTTTAGAGGTTTTTTTGTGTTAATAGTTTTTTTTTTTTTTTTTTTTAAAAATTTTA 4399

Qy       1114 GGTAATCACATTGTAGCTTGCTTTGCAACCTATATAAGAGCGCATAAAACTGAAAAAATG 1173
              || || |  ||  || || |||| ||     ||  |    |||  | ||       ||
Db       4400 AGTGATTAGTTTAGAGGTTTGTTTGGAATAATGGTATTATTTAATATTATTGGTTTTTTG 4459

Qy       1174 ACATTGGCATCATGTATAATGGCTCTCTTGACTATAACTACAATGTTGTTTGTTGCATAT 1233
              |||  |   | | || ||  | |  |     |||      | | |||| ||| || |
Db       4460 GAATTTGGGTATTTTTTAGTGTGTTTTATATTGTTAATATTTTTTTAGTTTTTTGGATTT 4519

Qy       1234 TT 1235
              ||
Db       4520 TT 4521


RESULT 13
US-10-311-455-197
; Sequence 197, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 197
;   LENGTH: 14924
;   TYPE: DNA
```

```
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 208..209, 211, 213, 227
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-197

  Query Match             3.5%;  Score 47.6;  DB 6;  Length 14924;
  Best Local Similarity   48.0%;  Pred. No. 12;
  Matches  166;  Conservative    0;  Mismatches  179;  Indels    1;  Gaps    1;

Qy        948 ACTGAGTAAATCGTTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATT 1007
              | | | ||| |  ||    |  | | | | |   | ||   ||||| | || || |  ||
Db       6176 ATTTAATAATTATTTTTTTAAAGTATTTAAGTGTTTATTGATTTTTTTTTTAACGGTTTT 6235

Qy       1008 AACTGGTGTATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCG 1067
              | |  | |  | |  | ||||| |||||| |   | |     |  |||   ||   | |
Db       6236 TATTTTTTTTAAGAGGTAATTATTGTTATTAATTATTTTGTATGTAAAATATGTTTATTG 6295

Qy       1068 GTTTTTGGGATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGT 1127
               |||    || | |    |    |||| |  | |   ||| || ||||  | | | |
Db       6296 AATTTATTTATAGAAAATAAATAATATTTTTTTGTAAAATATTTTTTGTAAAAAAAGTTT 6355

Qy       1128 AGCTTGCTTTGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATG 1187
                |    ||  |  | |||      | |||    | | | |||     ||  ||  |
Db       6356 TATATAAAATG-TATATGTATGTTTATGTATATGTTTTATATAATATATATGTTATATTA 6414

Qy       1188 TATAATGGCTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAG 1247
              ||||||   | | ||   ||||  || |      ||| ||   ||| || |  | |   |
Db       6415 TATAATATATATTTTTTTTATATATATATAGATTTTTTATGTATATATATAAATTTTG 6474

Qy       1248 GTTCTACATGTTAATGTATGCAGCACTAACGTGGTTATATTTTGTT 1293
                | || || || | || |    |||      |||||  | ||
Db       6475 TATATATATTTTTATATGTGTATATGTAAAAATAGTATATGGTATT 6520


RESULT 14
US-10-240-452-21
; Sequence 21, Application US/10240452
; Publication No. US20030162194A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Apoptosis
;  FILE REFERENCE: 5013.1006
;  CURRENT APPLICATION NUMBER: US/10/240,452
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03969
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 78
; SEQ ID NO 21
;   LENGTH: 14924
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (208..209, 211, 213, 227)
US-10-240-452-21

  Query Match             3.5%;  Score 47.6;  DB 6;  Length 14924;
  Best Local Similarity   48.0%;  Pred. No. 12;
  Matches  166;  Conservative    0;  Mismatches  179;  Indels    1;  Gaps    1;

Qy        948 ACTGAGTAAATCGTTCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATT 1007
              | | | ||| |  ||    |  | | | | |   | ||   ||||| | || || |  ||
Db       6176 ATTTAATAATTATTTTTTTAAAGTATTTAAGTGTTTATTGATTTTTTTTTTAACGGTTTT 6235

Qy       1008 AACTGGTGTATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCG 1067
              | |  | |  | |  | ||||| |||||| |   | |     |  |||   ||   | |
Db       6236 TATTTTTTTTAAGAGGTAATTATTGTTATTAATTATTTTGTATGTAAAATATGTTTATTG 6295

Qy       1068 GTTTTTGGGATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGT 1127
               |||    || | |    |    |||| |  | |   ||| || ||||  | | | |
Db       6296 AATTTATTTATAGAAAATAAATAATATTTTTTTGTAAAATATTTTTTGTAAAAAAAGTTT 6355

Qy       1128 AGCTTGCTTTGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATG 1187
                |    ||  |  | |||      | |||    | | | |||     ||  ||  |
Db       6356 TATATAAAATG-TATATGTATGTTTATGTATATGTTTTATATAATATATATGTTATATTA 6414

Qy       1188 TATAATGGCTCTCTTGACTATAACTACAATGTTGTTTGTTGCATATTTAGAGTACTCGAG 1247
              ||||||   | | ||   ||||  || |      ||| ||   ||| || |  | |   |
Db       6415 TATAATATATATTTTTTTTATATATATATAGATTTTTTATGTATATATATAAATTTTG 6474

Qy       1248 GTTCTACATGTTAATGTATGCAGCACTAACGTGGTTATATTTTGTT 1293
                | || || || | || |    |||      |||||  | ||
Db       6475 TATATATATTTTTATATGTGTATATGTAAAAATAGTATATGGTATT 6520


RESULT 15
US-10-311-455-1190
; Sequence 1190, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1190
;   LENGTH: 6191
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1190

  Query Match             3.5%;  Score 46.6;  DB 6;  Length 6191;
  Best Local Similarity   47.4%;  Pred. No. 13;
  Matches  139;  Conservative    0;  Mismatches  154;  Indels    0;  Gaps    0;

Qy        427 TTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCT 486
              || ||||||   | | |||    ||||  | |||||  || ||| | |  | | |  ||
Db       4845 TTTTTATTATTTAAATTTATGTTTTTTTTTTTTTTTTTTTATTTTTTTTTTATTAGTTTTG 4904

Qy        487 GAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCT 546
                | |  ||   |||  |     | |  | | ||| |   | | |       ||||  |
Db       4905 TTAAGTGGTTAATTATATTAATTTTTTTAAAGAATTAATTTTTTTTTAGTTTTGTTAATT 4964

Qy        547 GGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATA 606
              | |  ||  ||     ||  |  ||| |   ||| |  | | |  |   | |   |
Db       4965 TTTTTTATGTAAGATTATTTTTTATTTATTAATTTTTGTTTTTGTTTTTATTATTTTTTT 5024

Qy        607 GCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTC 666
```

```
                 | ||    |     ||    |     |   ||||||  | | | | ||||||||||| ||||
Db       5025 ATTTTTATTGTTTTTAGGTTTGATTTTTTATTTTTTTTTTTATTTTTTTAAAAAAGTTTTT 5084

Qy        667 CTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGT 719
                |       | || ||  | ||  | |  |    | |    | ||||||  |    ||
Db       5085 TTTTTTAATGTATAGATTTAAGGTTATAAATTTTATTTTAGTATTTTAGTAGT 5137


Search completed: May 12, 2006, 05:34:35
Job time : 862.623 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:52:36 ; Search time 683.239 Seconds
(without alignments)
8004.956 Million cell updates/sec

Title: US-09-423-093B-3_COPY_1019_2359
Perfect score: 1341
Sequence: 1 atgaatcgtattattagaat.....tcaagagttcttatgagtaa 1341

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/2/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/2/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	132.6	9.9	1353	18	US-11-079-463-635	Sequence 635, App
c 2	45.4	3.4	1071	10	US-10-793-626-1753	Sequence 1753, Ap
c 3	45.4	3.4	3053	10	US-10-793-626-3389	Sequence 3389, Ap
4	45.4	3.4	3159	10	US-10-793-626-4027	Sequence 4027, Ap
c 5	45.2	3.4	552	7	US-09-925-065A-456703	Sequence 456703,
c 6	45.2	3.4	553	12	US-10-301-480-514729	Sequence 514729,
c 7	45.2	3.4	553	12	US-10-301-480-1128138	Sequence 1128138,
8	44.2	3.3	173995	10	US-10-995-561-13277	Sequence 13277, A
c 9	43.2	3.2	1483	10	US-10-750-185-39052	Sequence 39052, A
c 10	43.2	3.2	1483	10	US-10-750-623-39052	Sequence 39052, A
11	42.4	3.2	90616	11	US-10-330-773-573	Sequence 573, App
c 12	41.8	3.1	608	7	US-09-925-065A-70061	Sequence 70061, A
c 13	41.8	3.1	608	11	US-10-301-480-171300	Sequence 171300,
c 14	41.8	3.1	608	12	US-10-301-480-784709	Sequence 784709,
c 15	41.8	3.1	1564	11	US-10-932-182A-166165	Sequence 166165,
c 16	41.8	3.1	1564	11	US-10-932-182A-166165	Sequence 166165,
17	41.8	3.1	1829	7	US-09-925-065A-71327	Sequence 71327, A
18	41.8	3.1	1829	7	US-09-925-065A-71329	Sequence 71329, A
19	41.8	3.1	1829	7	US-09-925-065A-71330	Sequence 71330, A
20	41.8	3.1	1829	11	US-10-301-480-172566	Sequence 172566,
21	41.8	3.1	1829	11	US-10-301-480-172568	Sequence 172568,
22	41.8	3.1	1829	11	US-10-301-480-172569	Sequence 172569,
23	41.8	3.1	1829	12	US-10-301-480-785975	Sequence 785975,
24	41.8	3.1	1829	12	US-10-301-480-785977	Sequence 785977,
25	41.8	3.1	1829	12	US-10-301-480-785978	Sequence 785978,
c 26	41.6	3.1	937	7	US-09-925-065A-713889	Sequence 713889,
c 27	41.6	3.1	231004	11	US-10-330-773-832	Sequence 832, App
28	41.4	3.1	627	7	US-09-925-065A-475658	Sequence 475658,
29	41.4	3.1	6656	10	US-10-240-708-75	Sequence 75, Appl
c 30	41.2	3.1	879	12	US-10-301-480-561830	Sequence 561830,
c 31	41.2	3.1	879	12	US-10-301-480-1175239	Sequence 1175239,
c 32	41	3.1	882	12	US-10-301-480-550973	Sequence 550973,
c 33	41	3.1	882	12	US-10-301-480-1164382	Sequence 1164382,
c 34	41	3.1	998	12	US-10-301-480-550974	Sequence 550974,
c 35	41	3.1	998	12	US-10-301-480-1164383	Sequence 1164383,
36	41	3.1	10619	10	US-10-240-708-3	Sequence 3, Appli
c 37	40.8	3.0	355	7	US-09-925-065A-89519	Sequence 89519, A
c 38	40.8	3.0	355	11	US-10-301-480-190760	Sequence 190760,
c 39	40.8	3.0	355	12	US-10-301-480-804169	Sequence 804169,
40	40.8	3.0	585	7	US-09-925-065A-608317	Sequence 608317,
41	40.8	3.0	990	12	US-10-301-480-593946	Sequence 593946,
42	40.8	3.0	990	12	US-10-301-480-1207355	Sequence 1207355,
c 43	40.8	3.0	1334	7	US-09-925-065A-55241	Sequence 55241, A
c 44	40.8	3.0	1334	7	US-09-925-065A-55242	Sequence 55242, A
c 45	40.8	3.0	1334	7	US-09-925-065A-55243	Sequence 55243, A

ALIGNMENTS

RESULT 1
US-11-079-463-635
; Sequence 635, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 635
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-635

Query Match 9.9%; Score 132.6; DB 18; Length 1353;
Best Local Similarity 46.9%; Pred. No. 1.3e-16;
Matches 559; Conservative 0; Mismatches 614; Indels 18; Gaps 4;

```
Qy          2 TGAATCGTATTATTAGAATGTTAGGTGTAGATAAAGCAATTCGTTATGTTATTTTTGGTA 61
              | ||   |||   ||   | ||     |||||   ||  |   |||   |   ||  |||
Db         17 TTAAAAATATATATAATTTTTTGCAATTAGATGCTGCTGTCATTTACTCTTCATTAAGTA 76

Qy         62 AGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAATATCACACCATTTATCTA 121
              ||||  |   ||  ||     || || ||      |     |||||  || |  | ||| |
Db         77 AGATCCTCAGTGGCTTTGGAGGCTTTCTAACGGTTTATTTAATCGCAAAAAAATTAACAT 136

Qy        122 AAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTTG 181
               ||  | ||| || || ||||||| ||||     ||| ||     |||||  || | ||||
Db        137 TAGTAGAACAAGGATACTATTATACTTTTATCAGTGTATTGTATATACAGGTATTTTTTG 196
```

```
Qy        182 AATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAAT 241
              ||||||||||     |  | |||| ||| || | |    || || |   ||   ||| | |
Db        197 AATTGGGGCTTAACTCTATTATCACTCAGTTTGTTGCTCACGAGAAAGCACATTTAGATT 256

Qy        242 ATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGT 301
                |      ||    ||| |  |||| ||       ||      ||||     ||
Db        257 GGAAAGGAAAGGA---TGATTTAGTAGGAAAAGAGTTTCATCTCTCTCGTTTAGCTTCTG 313

Qy        302 TATTTCGGTTGGCAATAAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTCG 361
              |||| || ||      | ||||  ||| |      | ||| |   | || |  |  | |
Db        314 TATTACGATTATGCGTTAAATATTATTCCTATTTGGCTATAGGATTACTGATTGTTTTGT 373

Qy        362 GTCCCATCGGGTATGTTTTTTTTACGCAAAAAGAAGGCTTAGGTGTACCTTGGCAAGGGG 421
              |      || |||||  ||||||  |    |||        | || ||   |||| ||
Db        374 TTATAGGTGGCTATGTCTTTTTTTCAATAAATAGTAATATTGGGGTTAGTTGGAAAATTC 433

Qy        422 CATGGTTATTATTAACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTG 481
              | |||||| |  |     |   |||  ||||     | ||  |   | || |  | ||
Db        434 CTTGGTTACTGCTTTGTGTTTCAACATCTTTATCGTTCTTCTTAAATCCTTTCTTATCCT 493

Qy        482 TCGCTGAAGGGAGTGGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGT 541
              ||   |||||        |||||  |  || ||     |   ||  |  |   ||||
Db        494 TCTTAGAAGGATTAAATTTAATGAAAGAGGTTTGTTTTATCCGTTTTATTCAGCAAACTG 553

Qy        542 TAGCTGGTATATTGGCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTG 601
              ||  |    |  |   |       ||| ||  | ||  |     |||||||  |
Db        554 TAAGTTTGCTGATATTATGGGTGGGACTGATAGGAGGTATGAAGCTATATGTTGG----- 608

Qy        602 CAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAA 661
                | | ||  ||       |  |  | |   |||| | || |  || | ||| | | |
Db        609 ---AGGATGTTCGAGTTTAGCTGGTGGTGTGGCTATTCTTATTTTTGTTTCTTACAGATA 665

Qy        662 TTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTA 721
                  |    | ||  ||    |||       || || |||      |  ||||   |
Db        666 CAGAATTTTGTTTTTGAATATATATGGGAAAGTA-ACTATTCATTTTATTAATTACAAGA 724

Qy        722 ATGAAATATTTCCTATGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTT 781
              | ||  ||||||||  | ||||||    | ||| |  | ||| |     | | |||  |||
Db        725 AAGAGATATTTCCATTTCAATGGAAGGTAGCTGTGGGATGGCTTAGCAGTTCTTTGGTTT 784

Qy        782 ATTTTGTTATGACCCCCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAG 841
              |    ||  | |  || |||   ||        | ||   |     ||| |||||  | |
Db        785 TTCAGTTTTTCAATCCGATTCTGTTTGCCACAATAGGTAGTGCTGCTGCTGGGCAACTTG 844

Qy        842 GGATGTCTTTAACATTATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACTA 901
              | ||| | ||||    |    ||   |||      ||       ||   ||||||| |||||
Db        845 GAATGACATTAAGTGTGATAAACGGGGTCTCATCTGTTTCAATGAACTGGATATACACTA 904

Qy        902 AATATCCAAAATGGGGAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCGT 961
              ||   || || | |      | |||| ||       | |  ||    ||| ||  |||||| |
Db        905 AAGTACCCAATTTGTCTAAATTGGTGTCATTACGAGATTTTAAAGAATTGGATAAATCAT 964

Qy        962 TCAAAAGTGCAGTAATGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATACA 1021
              |    |   | ||     ||      ||||    |  |  ||||| |    ||||
Db        965 TTTCTAAGATACTAGCTGTATTGGTGCTTTTGAGTTGTATGGGATTTATCATTGTAG--- 1021

Qy       1022 TTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTGC 1081
              | |||||  |  |  | ||    |       | ||||       |||  |     ||
Db       1022 ---CTTTATGTCTTATTCAGTTTAATATACTGCATATTGCGTCTAAGTTGCTTCCGATGT 1078

Qy       1082 AGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCAA 1141
              |   ||  |  |  | ||  ||| |   | |     |  || |||  | | || |  || |
Db       1079 CGTTGTTCCTAATAATGTCATTATCCTCTGTATTCACACAGATTACATCCTGTTGGGCTA 1138

Qy       1142 CCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAA 1192
                ||  | |||     ||| |  |||     |   |  || ||| |  || |
Db       1139 TTTACTTTAGATGCTTTAAGAAAGAACCTTTCTTAAGGGTATCTTTAATTA 1189


RESULT 2
US-10-793-626-1753/c
; Sequence 1753, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1753
;   LENGTH: 1071
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1753

  Query Match             3.4%;  Score 45.4;  DB 10;  Length 1071;
  Best Local Similarity   51.8%;  Pred. No. 19;
  Matches  103;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy       1017 ATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGG 1076
              |||  ||||||  |||| | | |    |||||    |||  |||         | |||  |
Db       1026 ATAGTTTTCATATTGGTCAGTAATTAAATCTGAACTAAATCTTGCTTCAATATCTTTCAG 967

Qy       1077 ATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTT 1136
                |    ||||   | |||| || |     |     ||| |||     |  |||| |  ||
Db        966 CATTTGTGATTGCATTTTTTGATATAACTCTGGATTTGATAAAAGTTTAATAGCATATTT 907

Qy       1137 TGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGC 1196
              |||| | | | ||  | | | || | ||   | ||   || |  | ||||||  | ||| |
Db        906 TGCAGCTTGTGTACTATCCCCTATATCTACTATAAATCCAGTTTCTTCATGTCTGATGAC 847

Qy       1197 TCTCTTGACTATAACTACA 1215
                 || |    | || ||
Db        846 CTCTTTAATACCACCTGCA 828


RESULT 3
US-10-793-626-3389/c
; Sequence 3389, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3389
;   LENGTH: 3053
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3389

  Query Match             3.4%;  Score 45.4;  DB 10;  Length 3053;
  Best Local Similarity   51.8%;  Pred. No. 24;
  Matches  103;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy       1017 ATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGG 1076
              |||  ||||||  |||| | | |    |||||    |||  |||         | |||  |
Db       2945 ATAGTTTTCATATTGGTCAGTAATTAAATCTGAACTAAATCTTGCTTCAATATCTTTCAG 2886

Qy       1077 ATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTT 1136
```

```
Db       2885 CATTTGTGATTGCATTTTTTGATATAACTCTGGATTTGATAAAAGTTTAATAGCATATTT 2826

Qy       1137 TGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGC 1196

Db       2825 TGCAGCTTGTGTACTATCCCCTATATCTACTATAAATCCAGTTTCTTCATGTCTGATGAC 2766

Qy       1197 TCTCTTGACTATAACTACA 1215

Db       2765 CTCTTTAATACCACCTGCA 2747


RESULT 4
US-10-793-626-4027
; Sequence 4027, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/10/793,626
;  CURRENT FILING DATE:  2004-03-04
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4027
;   LENGTH: 3159
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4027

  Query Match             3.4%;  Score 45.4;  DB 10;  Length 3159;
  Best Local Similarity   51.8%;  Pred. No. 24;
  Matches  103;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy       1017 ATACATTTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGG 1076

Db       2357 ATAGTTTTCATATTGGTCAGTAATTAAATCTGAACTAAATCTTGCTTCAATATCTTTCAG 2416

Qy       1077 ATTGCAGGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTT 1136

Db       2417 CATTTGTGATTGCATTTTTTGATATAACTCTGGATTTGATAAAAGTTTAATAGCATATTT 2476

Qy       1137 TGCAACCTATATAAGAGCGCATAAAACTGAAAAAATGACATTGGCATCATGTATAATGGC 1196

Db       2477 TGCAGCTTGTGTACTATCCCCTATATCTACTATAAATCCAGTTTCTTCATGTCTGATGAC 2536

Qy       1197 TCTCTTGACTATAACTACA 1215

Db       2537 CTCTTTAATACCACCTGCA 2555


RESULT 5
US-09-925-065A-456703/c
; Sequence 456703, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 456703
;   LENGTH: 552
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-456703

  Query Match             3.4%;  Score 45.2;  DB 7;  Length 552;
  Best Local Similarity   46.7%;  Pred. No. 18;
  Matches  140;  Conservative    1;  Mismatches  159;  Indels    0;  Gaps    0;

Qy       436 ACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGT 495

Db       501 AAAAAATTATCAAAATTTATTGTATAAATTCTTCCATTACTTTCAAAAATTCACCARTTT 442

Qy       496 GGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATATTG 555

Db       441 TAGTTATATCCATCACACTGAAGTTCCTATATGGCACTGCTGTTGTTGTTTTCCAGATAG 382

Qy       556 GCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTTCA 615

Db       381 TTTCTTCATTTGTTTACTAGTGTCTTTCTATTATGAACATCAAATTTCATAGCTTGGTGA 322

Qy       616 GGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCT 675

Db       321 AGAAGTTTCATGAGTTCCTTTTTGAATATGGTATAATAAAATATAAGGTCCTATGAATCT 262

Qy       676 TTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCT 735

Db       261 AAAAAAATATTAGTATAATATAAAAAGGGAGCAATAAAATTGGCTAATCAAGTACATCAT 202


RESULT 6
US-10-301-480-514729/c
; Sequence 514729, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 514729
;   LENGTH: 553
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-514729

  Query Match             3.4%;  Score 45.2;  DB 12;  Length 553;
  Best Local Similarity   46.7%;  Pred. No. 18;
  Matches  140;  Conservative    1;  Mismatches  159;  Indels    0;  Gaps    0;

Qy       436 ACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGT 495

Db       502 AAAAAATTATCAAAATTTATTGTATAAATTCTTCCATTACTTTCAAAAATTCACCARTTT 443

Qy       496 GGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATATTG 555

Db       442 TAGTTATATCCATCACACTGAAGTTCCTATATGGCACTGCTGTTGTTGTTTTCCAGATAG 383

Qy       556 GCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTTCA 615
```

```
Db        382 TTTCTTCATTTGTTTACTAGTGTCTTTCTATTATGAACATCAAATTTCATAGCTTGGTGA 323

Qy        616 GGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCT 675

Db        322 AGAAGTTTCATGAGTTCCTTTTTGAATATGGTATAATAAAATATAAGGTCCTATGAATCT 263

Qy        676 TTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCT 735

Db        262 AAAAAAATATTAGTATAATATAAAAAGGGAGCAATAAAATTGGCTAATCAAGTACATCAT 203


RESULT 7
US-10-301-480-1128138/c
; Sequence 1128138, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1128138
;   LENGTH: 553
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1128138

  Query Match             3.4%;  Score 45.2;  DB 12;  Length 553;
  Best Local Similarity   46.7%;  Pred. No. 18;
  Matches  140;  Conservative    1;  Mismatches  159;  Indels    0;  Gaps    0;

Qy        436 ACAATAGTTACAGCTTTTAATATTTTTCTTGTTTCTGTACTTTCTGTCGCTGAAGGGAGT 495

Db        502 AAAAAATTATCAAAATTTATTGTATAAATTCTTCCATTACTTTCAAAAATTCACCARTTT 443

Qy        496 GGGTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATATTG 555

Db        442 TAGTTATATCCATCACACTGAAGTTCCTATATGGCACTGCTGTTGTTGTTTTCCAGATAG 383

Qy        556 GCAGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTTCA 615

Db        382 TTTCTTCATTTGTTTACTAGTGTCTTTCTATTATGAACATCAAATTTCATAGCTTGGTGA 323

Qy        616 GGGACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCT 675

Db        322 AGAAGTTTCATGAGTTCCTTTTTGAATATGGTATAATAAAATATAAGGTCCTATGAATCT 263

Qy        676 TTAAAGCATAAAAATAAATATACTGAAGGTGGTATTTCATGGGTTAATGAAATATTTCCT 735

Db        262 AAAAAAATATTAGTATAATATAAAAAGGGAGCAATAAAATTGGCTAATCAAGTACATCAT 203


RESULT 8
US-10-995-561-13277
; Sequence 13277, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13277
;   LENGTH: 173995
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-995-561-13277

  Query Match             3.3%;  Score 44.2;  DB 10;  Length 173995;
  Best Local Similarity   48.2%;  Pred. No. 96;
  Matches  124;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy        138 TTATTATACATTTAATTCAGTAGTGGCACTACAGATAATATTTGAATTGGGGCTATCAAC 197

Db     152253 TTTGTAAAAAGCTATTGTAGTTGTTATCTTAAAAGAAATGAGTATATATGTTCTTTAGAA 152312

Qy        198 GGTAATCATTCAATTCGCTAGCCATGAAATGTCAGCGTTAAAATATGATTATTCTGAACG 257

Db     152313 GTCTTTAATTAAATAAGTTTCTCATTAAAGTAAAAAGTTAAAATAAGAAGGATGCAATAG 152372

Qy        258 AGATATTATAGGTGAAAGTAAAAATAAGCAACGTTACCTATCGTTATTTCGGTTGGCAAT 317

Db     152373 AAATATCCTGGAAGCAGAGAAGAATTAGAAATAACAATTATAAATAAAAATATCAAGTAT 152432

Qy        318 AAAATGGTATGCAGTAATAGCTTTGCTAATAATATTAATAGTCGGTCCCATCGGGTATGT 377

Db     152433 TGGTAGTAATTTGCTGATTGCTTTTTAAATATTAATATTGTCCACCTCAATAGGAAATGC 152492

Qy        378 TTTTTTTACGCAAAAAG 394

Db     152493 TGCTCTTATTACCAAAG 152509


RESULT 9
US-10-750-185-39052/c
; Sequence 39052, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-2
;  CURRENT APPLICATION NUMBER: US/10/750,185
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 39052
;   LENGTH: 1483
;   TYPE: DNA
;   ORGANISM: Bovine  19866881927317
US-10-750-185-39052

  Query Match             3.2%;  Score 43.2;  DB 10;  Length 1483;
  Best Local Similarity   48.8%;  Pred. No. 56;
  Matches  117;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy         45 TTATGTTATTTTTGGTAAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAAT 104

Db        846 TTATATTTATACTTAAATAAATTTATATTTATTAATATATTTACATTTTATAATTATATA 787

Qy        105 ATCACACCATTTATCTAAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGC 164

Db        786 TTAATATAATTTATATTTAGATTTATATTTAAATATATTTACAAATATGTAATTGATATA 727

Qy        165 ACTACAGATAATATTTGAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGA 224

Db        726 TCTTTATGATATATCTATATTATATTTATCAACTATAATTTATATTTTAAATATAGCTTA 667
```

```
Qy        225 AATGTCAGCGTTAAAATATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAA 284
              || |         | |||| || ||  | |      ||||  |    |   | | | | | |
Db        666 TATTTATATAACTATATATAATAATGTTTATATTTATATATTTATTTGCATTTATATTTA 607


RESULT 10
US-10-750-623-39052/c
; Sequence 39052, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
;  APPLICANT: MMI GENOMICS, INC.
;  APPLICANT:  DeNISE, Sue K.
;  APPLICANT:  KERR, Richard
;  APPLICANT:  ROSENFELD, David
;  APPLICANT:  HOLM, Tom
;  APPLICANT:  BATES, Stephen
;  APPLICANT:  FANTIN, Dennis
;  TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
;  FILE REFERENCE: MMI1100-1
;  CURRENT APPLICATION NUMBER: US/10/750,623
;  CURRENT FILING DATE:  2003-12-31
;  PRIOR APPLICATION NUMBER: US 60/437,482
;  PRIOR FILING DATE: 2002-12-31
;  NUMBER OF SEQ ID NOS: 64922
;  SOFTWARE: PatentIN version 3.1
; SEQ ID NO 39052
;   LENGTH: 1483
;   TYPE: DNA
;   ORGANISM: Bovine  19866881927317
US-10-750-623-39052

  Query Match             3.2%;  Score 43.2;  DB 10;  Length 1483;
  Best Local Similarity   48.8%;  Pred. No. 56;
  Matches  117;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy         45 TTATGTTATTTTTGGTAAGATAATATCTGTATTAACGGGTTTACTGTTAATAATGTTAAT 104
              |||| ||  |  |   |  |   ||| | ||||||    |||||  || |||||  ||
Db        846 TTATATTTATACTTAAATAAATTTATATTTATTAATATATTTACATTTTATAATTATATA 787

Qy        105 ATCACACCATTTATCTAAAGACGCACAGGGCTATTATTATACATTTAATTCAGTAGTGGC 164
              | | |  |||||| |  |||   | |     ||   | ||||  ||  | | |  |
Db        786 TTAATATAATTTATATTTAGATTTATATTTAAATATATTTACAAATATGTAATTGATATA 727

Qy        165 ACTACAGATAATATTTGAATTGGGGCTATCAACGGTAATCATTCAATTCGCTAGCCATGA 224
              ||  |    |||| |  |||     |||||||  ||||   |   ||   ||    | |
Db        726 TCTTTATGATATATCTATATTATATTTATCAACTATAATTTATATTTTAAATATAGCTTA 667

Qy        225 AATGTCAGCGTTAAAATATGATTATTCTGAACGAGATATTATAGGTGAAAGTAAAAATAA 284
              || |         | |||| || ||  | |      ||||  |    |   | | | | | |
Db        666 TATTTATATAACTATATATAATAATGTTTATATTTATATATTTATTTGCATTTATATTTA 607


RESULT 11
US-10-330-773-573
; Sequence 573, Application US/10330773
; Publication No. US20060040262A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Compositions and Methods in Cancer
;  FILE REFERENCE: 529452001300
;  CURRENT APPLICATION NUMBER: US/10/330,773
;  CURRENT FILING DATE:  2002-12-27
;  NUMBER OF SEQ ID NOS: 981
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 573
;   LENGTH: 90616
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(90616)
;   OTHER INFORMATION: n = A,T,C or G
US-10-330-773-573

  Query Match             3.2%;  Score 42.4;  DB 11;  Length 90616;
  Best Local Similarity   56.4%;  Pred. No. 1.9e+02;
  Matches   79;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy       1023 TTCATTATGGTTATTGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTGCA 1082
              || | | |||   ||    || | |  || | | |||    |    |||||   |
Db      14349 TTGACTTTGGGATTTCTTTTTTTTTTTTTATAGATTTTAAATGTTTTTTTTTAAAGATT 14408

Qy       1083 GGATTTTTTCTTTTTATCTTTAGCAATTATTGGTAATCACATTGTAGCTTGCTTTGCAAC 1142
                || ||| ||| | | ||||   |||||  ||||  ||||||||||  |||   |
Db      14409 TATTTATTTATTTATTTATTTATTTATTATATGTAAGTACATTGTAGCTGTCTTCAGACA 14468

Qy       1143 CTATATAAGAGCGCATAAAA 1162
              ||  | |||||  |||| | |
Db      14469 CTCCAGAAGAGGGCATCAGA 14488


RESULT 12
US-09-925-065A-70061/c
; Sequence 70061, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70061
;   LENGTH: 608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-70061

  Query Match             3.1%;  Score 41.8;  DB 7;  Length 608;
  Best Local Similarity   54.1%;  Pred. No. 88;
  Matches   85;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        588 ATATGCTACGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAA 647
              | |||     | |  |  || | |  |||  |  || || |  |  |||||| |||
Db        378 AAATGAGTATTTTTAAGGAGGTCTCACAGCTAACATTATTTCATAAATATTTTTATAATT 319

Qy        648 GTATTTTAAAAAAATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGG 707
              || || |||| |||      |  |   ||| | | |||| |  || | || | ||
Db        318 GTTTTGTAAATAAACAGCATTATATATTTTGAGGTATAACAGGAACTTTATTAAACAATT 259

Qy        708 TATTTCATGGGTTAATGAAATATTTCCTATGCAATGG 744
              || | ||   | ||| || ||||  |||| |  |||
Db        258 AATATTATATATAAATTAACTATTCTCTATTCTCTGG 222


RESULT 13
US-10-301-480-171300/c
; Sequence 171300, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
```

```
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 171300
;   LENGTH: 608
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-171300

  Query Match             3.1%;  Score 41.8;  DB 11;  Length 608;
  Best Local Similarity   54.1%;  Pred. No. 88;
  Matches   85;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        588 ATATGCTACGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAA 647
              | |||     | |  |  || | |  |||  |  || ||  |  |  ||||||| |||
Db        378 AAATGAGTATTTTTAAGGAGGTCTCACAGCTAACATTATTTCATAAATATTTTTATAATT 319

Qy        648 GTATTTTAAAAAAATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGG 707
              || || |||| |||       |  |   ||| | | |||| |  || | || | ||
Db        318 GTTTTGTAAATAAACAGCATTATATATTTTGAGGTATAACAGGAACTTTATTAAACAATT 259

Qy        708 TATTTCATGGGTTAATGAAATATTTCCTATGCAATGG 744
              || | ||   | ||| || ||||  |||| |  |||
Db        258 AATATTATATATAAATTAACTATTCTCTATTCTCTGG 222


RESULT 14
US-10-301-480-784709/c
; Sequence 784709, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 784709
;   LENGTH: 608
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-784709

  Query Match             3.1%;  Score 41.8;  DB 12;  Length 608;
  Best Local Similarity   54.1%;  Pred. No. 88;
  Matches   85;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy        588 ATATGCTACGTCTGCAATAGCTATTTCAGGGACTATCATATTCTCCATATTTTCATATAA 647
              | |||     | |  |  || | |  |||  |  || ||  |  |  ||||||| |||
Db        378 AAATGAGTATTTTTAAGGAGGTCTCACAGCTAACATTATTTCATAAATATTTTTATAATT 319

Qy        648 GTATTTTAAAAAAATTTTCCTGCAATCTTTAAAGCATAAAAATAAATATACTGAAGGTGG 707
              || || |||| |||       |  |   ||| | | |||| |  || | || | ||
Db        318 GTTTTGTAAATAAACAGCATTATATATTTTGAGGTATAACAGGAACTTTATTAAACAATT 259

Qy        708 TATTTCATGGGTTAATGAAATATTTCCTATGCAATGG 744
              || | ||   | ||| || ||||  |||| |  |||
Db        258 AATATTATATATAAATTAACTATTCTCTATTCTCTGG 222


RESULT 15
US-10-932-182A-166165/c
; Sequence 166165, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAO, YOSHIHIRO
;  APPLICANT:  NAKAMURA, NORIHISA
;  APPLICANT:  KODAMA, YUKIKO
;  APPLICANT:  FUJIMURA, TOMOKO
;  APPLICANT:  ASHIKARI, TOSHIHIKO
;  TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
;  FILE REFERENCE: 030685-043
;  CURRENT APPLICATION NUMBER: US/10/932,182A
;  CURRENT FILING DATE:  2004-09-02
;  NUMBER OF SEQ ID NOS: 197023
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166165
;   LENGTH: 1564
;   TYPE: DNA
;   ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166165

  Query Match             3.1%;  Score 41.8;  DB 11;  Length 1564;
  Best Local Similarity   42.8%;  Pred. No. 1.1e+02;
  Matches  264;  Conservative    0;  Mismatches  352;  Indels    1;  Gaps    1;

Qy        498 GTTAATTACTGATGTGAATAAAATGAGAATGTATCAGTCGCTGTTAGCTGGTATATTGGC 557
              ||| || |    | |  | |  || | |||  ||        ||||    |  ||||||
Db       1011 GTTTATAATCATTATTTAAATTATTATAATAAATATTATTATGTTTATTTATATATTTTA 952

Qy        558 AGTAAGCTTACTTATTAGTGGCTTTGGACTATATGCTACGTCTGCAATAGCTATTTCAGG 617
                |    | ||||   | ||| |  | |         |    | | |||  ||||    |
Db        951 TTATATTAGATTTATCTATAGCTATATATTTATCTAAATAAATACCATAAATATTATTAG 892

Qy        618 GACTATCATATTCTCCATATTTTCATATAAGTATTTTAAAAAAATTTTCCTGCAATCTTT 677
              | || |||||| |  |  ||    |||||   |  |  |  |  ||     ||
Db        891 TAATAGCATATTTTAATTTATTAGTAATAAGATCTAAAGTATCAAATTTAGATAAAGAAA 832

Qy        678 AAAGCATAAAAATAAATATACTGAAGGT-GGTATTTCATGGGTTAATGAAATATTTCCTA 736
              ||  | | ||| ||     || |  | || |||| ||   ||  |||  |||| | | | |
Db        831 AATTCTTTAAATTACCACTATTATATGTAGGTAATTTTAGGTATAAATAAATTTGTTCAA 772

Qy        737 TGCAATGGCGAATTGCTCTAAGTTGGATGTCAGGGTATTTTATTTATTTTGTTATGACCC 796
              | | ||    |||   ||||        | |||    |  | ||||| |||| |
Db        771 TACTATATAAAATATATCTATCAATACTTTCATTATCATATATTTTTTTTATATCATATA 712

Qy        797 CCATTGCATTCAAATATTTCGGGGCTATATATGCAGGGCAGTTAGGGATGTCTTTAACAT 856
              | |  | ||| |||   |        || || |   |    ||||   |  |   |
Db        711 CAAAAGTATTTAAACCATATTTTTTTACATCTAATGCAATATTAGAACTACCACGTAATC 652

Qy        857 TATGCAATATGGTAATGGCTACGGGCCTGGCTTGGATATCCACTAAATATCCAAAATGGG 916
              |||  ||      || |   |          |   |||   ||     | |||| ||
Db        651 TATATAAATGATTATTAAATCTATCATATATATTATTATATCCTGCTGAACCAATATAAA 592

Qy        917 GAGTAATGGTTTCCAACAAACAGCTTGCGGAACTGAGTAAATCGTTCAAAAGTGCAGTAA 976
                || | ||    |     ||   |     || |   |    |  || | |   |  |
Db        591 ATTTATTTGTCATTATATTACCAATATAATAAATACCTTGTTTTTTTATTAAATAATTTA 532

Qy        977 TGCAATCATCCTTTTTTGTCTTGACAGGATTAACTGGTGTATACATTTCATTATGGTTAT 1036
              |    || || ||||  |              |  |  |  |||  ||  |   |  | |
Db        531 TAATTTCTTCTTTTTTATTTGGACATTCTAAATTTACTCAATATTTTATAGATTTAATTT 472

Qy       1037 TGAAATTATCTGGTTCAAACATTGGCGAGCGGTTTTTGGGATTGCAGGATTTTTTCTTTT 1096
              |   ||| | |  |  ||  ||    |  | ||   ||     || || | || |
Db        471 TATGATTTTTTTTTATAATTCTTCTACACTTAGTATTTTTTTTTATTTATATTATATTCT 412

Qy       1097 TATCTTTAGCAATTATT 1113
              || |  |  || ||||
Db        411 AATTTAGAAGAAGTATT 395


Search completed: May 12, 2006, 06:55:05
Job time : 684.239 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 08:31:04 ; Search time 4319.22 Seconds
(without alignments)
15792.702 Million cell updates/sec

Title: US-09-423-093B-3_COPY_5114_6313
Perfect score: 1200
Sequence: 1 atgctgccatttccaccagg.....atagacggtgcacgcgataa 1200

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1200	100.0	12441	1	SERFBC2	X61917 S.enterica
	2	1200	100.0	12441	6	BD085556	BD085556 Nucleic a
	3	1200	100.0	12446	6	AX029263	AX029263 Sequence
c	4	68.8	5.7	194784	5	BX890608	BX890608 Zebrafish
	5	65.2	5.4	182150	5	BX899180	BX899180 Zebrafish
	6	63.8	5.3	1141	6	AR579680	AR579680 Sequence
	7	63.8	5.3	1141	6	AX083744	AX083744 Sequence
c	8	63.2	5.3	331039	2	AC116979	AC116979 Dictyoste
	9	62.8	5.2	155885	14	CR848672	CR848672 Danio rer
c	10	61.4	5.1	157210	5	BX928756	BX928756 Zebrafish
c	11	60.4	5.0	142687	14	CR388190	CR388190 Danio rer
c	12	60.2	5.0	202836	14	CR847804	CR847804 Danio rer
c	13	59.6	5.0	202208	14	CR759891	CR759891 Danio rer
	14	59	4.9	181792	8	AC098822	AC098822 Homo sapi
	15	58.6	4.9	18154	6	AX345156	AX345156 Sequence
	16	58.4	4.9	97766	5	CR352306	CR352306 Zebrafish
	17	57.8	4.8	110000	2	AC116305_1	Continuation (2 of
	18	57.8	4.8	156236	5	BX088709	BX088709 Zebrafish
c	19	57.8	4.8	349751	2	PFMAL4P3	AL035476 Plasmodiu
c	20	57.4	4.8	157544	5	CR762435	CR762435 Zebrafish
	21	56.8	4.7	173048	5	AL772154	AL772154 Zebrafish
c	22	56.4	4.7	18168	2	PF13PB1	AL929365 Plasmodiu
	23	56.4	4.7	107132	5	BX255941	BX255941 Zebrafish
	24	56.4	4.7	204652	14	PFMAL13P6	AL049183 Plasmodiu
	25	56.4	4.7	206784	8	HS93L7	AL022401 Human DNA
c	26	56.2	4.7	145992	8	AC134919	AC134919 Homo sapi
c	27	56	4.7	115990	5	BX890614	BX890614 Zebrafish
c	28	55.8	4.7	341050	2	PFA929357	AL929357 Plasmodiu
c	29	55.6	4.6	43872	2	AY043293	AY043293 Tribolium
	30	55.6	4.6	228170	5	BX511086	BX511086 Zebrafish
	31	55.4	4.6	217722	5	BX284621	BX284621 Zebrafish
c	32	55.2	4.6	110000	2	AC116957_1	Continuation (2 of
c	33	55	4.6	83110	14	PFMAL13PB	AL109814 Plasmodiu
	34	55	4.6	110000	14	PFMAL13_13	Continuation (14 o
	35	55	4.6	159618	8	AC096750	AC096750 Homo sapi
c	36	55	4.6	182150	5	BX899180	BX899180 Zebrafish
	37	54.8	4.6	1916	2	AF503868	AF503868 Plasmodiu
c	38	54.8	4.6	110000	14	PFMAL7P1_10	Continuation (11 o
c	39	54.6	4.5	108767	5	CR735141	CR735141 Zebrafish
	40	54.6	4.5	137889	8	AC073269	AC073269 Homo sapi
c	41	54.6	4.5	176147	5	AL929105	AL929105 Zebrafish
	42	54.6	4.5	192929	14	AC005505	AC005505 Plasmodiu
	43	54.6	4.5	226847	14	BX664721	BX664721 Danio rer
	44	54.6	4.5	250713	2	AE014850	AE014850 Plasmodiu
c	45	54.6	4.5	256172	14	AC005139	AC005139 Plasmodiu

ALIGNMENTS

RESULT 1
SERFBC2
LOCUS SERFBC2 12441 bp DNA linear BCT 09-SEP-2004
DEFINITION S.enterica rfbJ gene cluster.
ACCESSION X61917
VERSION X61917.1 GI:47004
KEYWORDS abe gene; abequose synthase; galactosyltransferase; manB gene; manC gene; mannose-1-phosphate guanyltransferase; phosphomannomutase; wbaL gene; wbaP gene; wbaQ gene; wbaR gene; wbaW gene; wbaZ gene; wzx gene; wzy gene.
SOURCE Salmonella enterica
ORGANISM Salmonella enterica
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 12441)
AUTHORS Brown,P.K., Romana,L.K. and Reeves,P.R.
TITLE Molecular analysis of the rfb gene cluster of Salmonella serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B
JOURNAL Mol. Microbiol. 6 (10), 1385-1394 (1992)
PUBMED 1379320
REFERENCE 2
AUTHORS Reeves,P.
TITLE Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale
JOURNAL Trends Genet. 9 (1), 17-22 (1993)
PUBMED 8434412
REFERENCE 3
AUTHORS Liu,D., Haase,A.M., Lindqvist,L., Lindberg,A.A. and Reeves,P.R.
TITLE Glycosyl transferases of O-antigen biosynthesis in Salmonella enterica: identification and characterization of transferase genes of groups B, C2, and E1
JOURNAL J. Bacteriol. 175 (11), 3408-3413 (1993)
PUBMED 7684736
REFERENCE 4 (bases 1 to 12441)
AUTHORS Brown,P.K.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1991) P.K. Brown, University of Sydney, Dept of Microbiology, Maze Crescent, Sydney N S W 2006, AUSTRALIA
FEATURES Location/Qualifiers
source 1. .12441

```
                /organism="Salmonella enterica"
                /mol_type="genomic DNA"
                /strain="group C2, serovar muenchen M67"
                /db_xref="taxon:28901"
gene            138. .1019
                /gene="abe"
CDS             138. .1019
                /gene="abe"
                /codon_start=1
                /transl_table=11
                /product="abequose synthase"
                /protein_id="CAA43918.1"
                /db_xref="GI:47005"
                /db_xref="GOA:Q00329"
                /db_xref="UniProt/Swiss-Prot:Q00329"
                /translation="MLDVNKKILMTGATSFVGTHLLHSLIKEGYSIIALKRPITEPTI
                INTLIEWLNIQDIEKICQSSMNIHAIVHIATDYGRNRTPISEQYKCNVLLPTRLLELM
                PALKTKFFISTDSFFGKYEKHYGYMRSYMASKRHFVELSKIYVEEHPDVCFINLRLEH
                VYGERDKAGKIIPYVIKKMKNNEDIDCTIARQKRDFIYIDDVVSAYLKILKEGFNAGH
                YDVEVGTGKSIELKEVFEIIKKETHSSSKINYGAVAMRDDEIMESHANTSFLTRLGWS
                AEFSIEKGVKKMLSMKE"
CDS             1019. .2359
                /note="wzx
                Referred to as orf12.6 in reference [2]"
                /citation=[2]
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein with 12 predicted
                transmembrane segments which possibly functions in export
                of O-antigen"
                /protein_id="CAA43919.1"
                /db_xref="GI:47006"
                /db_xref="UniProt/TrEMBL:Q00339"
                /translation="MNRIIRMLGVDKAIRYVIFGKIISVLTGLLLIMLISHHLSKDAQ
                GYYYTFNSVVALQIIFELGLSTVIIQFASHEMSALKYDYSERDIIGESKNKQRYLSLF
                RLAIKWYAVIALLIILIVGPIGYVFFTQKEGLGVPWQGAWLLLTIVTAFNIFLVSVLS
                VAEGSGLITDVNKMRMYQSLLAGILAVSLLISGFGLYATSAIAISGTIIFSIFSYKYF
                KKIFLQSLKHKNKYTEGGISWVNEIFPMQWRIALSWMSGYFIYFVMTPIAFKYFGAIY
                AGQLGMSLTLCNMVMATGLAWISTKYPKWGVMVSNKQLAELSKSFKSAVMQSSFFVLT
                GLTGVYISLWLLKLSGSNIGERFLGLQDFFFLSLAIIGNHIVACFATYIRAHKTEKMT
                LASCIMALLTITTMLFVAYLEYSRFYMLMYAALTWLYFVPQTYIIFKRFKSSYE"
CDS             2352. .3314
                /note="unnamed protein product; wbaR"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA43909.1"
                /db_xref="GI:47007"
                /db_xref="UniProt/TrEMBL:Q00337"
                /translation="MSKKPLLTIAIPTYNRSSCLARLLDSIIQQENYCHDELEVIVCD
                NASTDETARIAKSGLDKIRNSTYHLNEENLGMDGNFQKCFELSNGKYLWMIGDDDLIV
                KNGISKVFSILKSRPALDMVYVNSAAKTELNYNADVRTSFYTNDVDFISDVKVMFTFI
                SGMICKKTDAIVKAVGIFSPQTTGKYLMHLTWQLPLLKQGGEFAVIHNNIIEAEPDNS
                GGYHLYKVFSNNLATIFDVFYPREHRVSKRVRASACLFLLNFIGDEDKTKNFATNNYL
                RDCDSAFIDLIIYKYGLRFFYLYPKTVPLFRKIKYIIKTVLMRK"
CDS             3361. .3975
                /note="unnamed protein product; wbaL
                orf14.9 possibly encodes an acetyltransferase"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA43910.1"
                /db_xref="GI:47008"
                /db_xref="UniProt/TrEMBL:Q00480"
                /translation="MFVYSLRLKLNLIISLLSKVRRKSKAKFLVLLSGYDFKMVGKNF
                KLNVKPYSAKNNTSSKWGSMRVGDNCWIEAVYNYGDEKFEPYLYIGDRICLSDNVHIS
                CVSCLILENDILIGSKVYIGDHSHGSYKVCSPKIEPPANKPLGDIAPIKIGNCCWIGD
                NAVILAGSEICDGCVIAANSVVKDLKVDKPCLIGGVPAKVIKVF"
CDS             3977. .5020
                /note="wbaQ
                Referred to as orf15.6 in reference [2]"
                /citation=[3]
                /codon_start=1
                /transl_table=11
                /product="rhamnosyl transferase"
                /protein_id="CAA43911.1"
                /db_xref="GI:47009"
                /db_xref="GOA:Q00482"
                /db_xref="UniProt/TrEMBL:Q00482"
                /translation="MNVFISICIPSYNRAEFLEPLLDSIYNQDYCLKNNDFEVIVCED
                KSPQRDEINSIIENYKAKNNKQNLYVNFNEDNLGYDKNLKKCISLTTGKYCMIMGNDD
                LLADGALSKIVKVLKANPEIVLATRAYGWFKENPNELCDTVRHLTDDTLFQPGADAIK
                FFFRRVGVISGFIVNAEKAKKLSSDLFDGRLYYQMYLAGMLMAEGQGYYFSDVMTLSR
                DTEAPDFGNAGTEKGVFTPGGYKPEGRIHMVEGLLLIAKYIEDTTKIDGVYAGIRKDL
                ANYFYPYIRDQLDLPLYTYIKMINKFRKMGFSNEKLFYVHAFLGYVLKRRGYDALIKY
                IRSKKGGTPRLGI"
gene            5114. .6313
                /gene="wzy"
CDS             5114. .6313
                /gene="wzy"
                /codon_start=1
                /transl_table=11
                /product="O antigen polymerase"
                /protein_id="CAA43912.1"
                /db_xref="GI:47010"
                /db_xref="GOA:Q00474"
                /db_xref="UniProt/Swiss-Prot:Q00474"
                /translation="MLPFPPGAILRDVLNVFFVALVLVRFVIDRKKTYFPLVFTIFSW
                SAVILWVIALTIFSPDKIQAIMGGRSYILFPAVFIALVILKVSYPQSLNIEKIVCYII
                FLMFMVATISIIDVLMNGEFIKLLGYDEHYAGEQLNLINSYDGMVRATGGFSDALNFG
                YMLTLGVLLCMECFSQGYKRLLMLIISFVLFIAICMSLTRGAILVAALIYALYIISNR
                KMLFCGITLFVIIIPVLAISTNIFDNYTEILIGRFTDSSQASRGSTQGRIDMAINSLN
                FLSEHPSGIGLGTQGSGNMLSVKDNRLNTDNYFFWIALETGIIGLIINIIYLASQFYS
                STLLNRIYGSHCSNMHYRLYFLFGSIYFISAALSSAPSSSTFSIYYWTVLALIPFLKL
                TNRRCTR"
CDS             6313. .7323
                /note="wbaW
                Referred to as orf17.9 in reference [2]"
                /citation=[3]
                /codon_start=1
                /transl_table=11
                /product="second mannosyl transferase"
                /protein_id="CAA43913.1"
                /db_xref="GI:47011"
                /db_xref="GOA:Q00481"
                /db_xref="UniProt/TrEMBL:Q00481"
                /translation="MNNKKVLMDISWSNKGGIGRFTDEISKLLCDISKEELYRKCASP
                LAPLGLAVNIFLRKKTDVVFLPGYIPPLFCSKKFIITIHDLNHLDLNDNSSLFKRLFY
                NFIIKRGCRKAYKIFTVSNFSKERIVAWSGVNPNKIVTVYNGVSSLFNADVKPLNLGY
                KYLLCVGNRKTHKNEKCVISAFAKADIDPSIKLVFTGNPCNDLEKLIIQHGLSERVKF
                FGFVSEKDLPSLYKGSLGLVFPSLYEGFGLPVVEGMACGIPVLTSLTSSLPEVAGDAA
                ILVDPLSEDAITKGISRLINDSELRKHLIQKGLLRAKRFNWQNVVSEIEMVLTEACDG
                NK"
CDS             7310. .8467
                /note="wbaZ
                Referred to as orf18.9 in reference [2]"
                /citation=[3]
                /codon_start=1
                /transl_table=11
                /product="first mannosyl transferase"
                /protein_id="CAA43914.1"
                /db_xref="GI:47012"
                /db_xref="GOA:Q00483"
                /db_xref="UniProt/TrEMBL:Q00483"
                /translation="MEINEIKISLVHEWLLSYAGSEQVSSAILHVFPEAKLYSVVDFL
                TDEQRRHFLGKYATTTFIQNLPKAKKFYQKYLPLMPLAIEQLDLSDANIIISSAHSVA
                KGVISGPDQLHISYVHSPIRYAWDLQHQYLNESNLNKGIKGWLAKWLLHKIRIWDSRT
                ANGVDHFIANSQYIARRIKKVYRREASVIYPPVDVDNFEVKNEKQDYYFTASRMVPYK
                RIDLIVEAFSKMPEKKLVVIGDGPEMKKIKSKATDNIKLLGYQSFPVLKEYMQSARAF
                VFAAEEDFGIIPVEAQACGTPVIAFGKGGALETVRPLGVEEEPTGIFFKEQNIASLHEA
                VSEFEKNASFFTSQACRKNAEKFSRSRFEQEFKNFVNEKWNLFKTEQIIKR"
gene            8482. .9903
                /gene="manC"
CDS             8482. .9903
                /gene="manC"
                /codon_start=1
                /transl_table=11
                /product="mannose-1-phosphate guanyltransferase"
```

```
                     /protein_id="CAA43915.1"
                     /db_xref="GI:47013"
                     /db_xref="GOA:Q00473"
                     /db_xref="UniProt/Swiss-Prot:Q00473"
                     /translation="MSKLIPVIMAGGIGSRLWPLSREEHPKQFLSVDGELSMLQNTIK
                     RLTPLLAGEPLVICNDSHRFLVAEQLRAINKLANNIILEPVGRNTAPAIALAAFCSLQ
                     NVVDEDPLLLVLAADHVIRDEKVFLKAINHAEFFATQGKLVTFGIVPTQAETGYGYIC
                     RGEAIGEDAFSVAEFVEKPDFDTARHYVESEKYYWNSGMFLFRASSYLQELKDLSPDI
                     YQACENAVGSINPDLDFIRIDKEAFAMCPSDSIDYAVMEHTRHAVVVPMNAGWSDVGS
                     WSSLWDISKKDPQRNVLHGDIFAYNSKDNYIYSEKSFISTIGVNNLVIVQTADALLVS
                     DKDSVQDVKKVVDYLKANNRNEHKKHLEVFRPWGKFSVIHSGDNYLVKRITVKPGAKF
                     AAQMHLHRAEHWIVVSGTACITKGEEIFTISENESTFIPANTVHTLKNPATIPLELIE
                     IQSGTYLAEDDIIRLEKHSGYLE"
     gene            9911. .11347
                     /gene="manB"

  Query Match           100.0%;  Score 1200;  DB 1;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 7.1e-197;
  Matches 1200;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCTGCCATTTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5114 ATGCTGCCATTTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCG 5173

Qy         61 TTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5174 TTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACT 5233

Qy        121 ATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5234 ATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGAT 5293

Qy        181 AAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5294 AAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCA 5353

Qy        241 TTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5354 TTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTAC 5413

Qy        301 ATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5414 ATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGA 5473

Qy        361 GAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5474 GAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATT 5533

Qy        421 AATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5534 AATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGA 5593

Qy        481 TATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5594 TATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGA 5653

Qy        541 TTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5654 TTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGA 5713

Qy        601 GGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5714 GGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATG 5773

Qy        661 CTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5774 CTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAAT 5833

Qy        721 ATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5834 ATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCG 5893

Qy        781 CGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAGAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5894 CGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAGAA 5953

Qy        841 CATCCATCAGGTATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5954 CATCCATCAGGTATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGAT 6013

Qy        901 AATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6014 AATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGC 6073

Qy        961 TTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6074 TTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGA 6133

Qy       1021 ATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6134 ATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATA 6193

Qy       1081 TATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6194 TATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTAT 6253

Qy       1141 TGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGACGGTGCACGCGATAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6254 TGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGACGGTGCACGCGATAA 6313


RESULT 2
BD085556
LOCUS       BD085556               12441 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085556
VERSION     BD085556.1  GI:22631166
KEYWORDS    JP 2001522248-A/3.
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 12441)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 3 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Salmonella enterica
            PN   JP 2001522248-A/3
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .12441
            FT                   /organism='Salmonella enterica'.
FEATURES             Location/Qualifiers
     source          1. .12441
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match           100.0%;  Score 1200;  DB 6;  Length 12441;
  Best Local Similarity  100.0%;  Pred. No. 7.1e-197;
  Matches 1200;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCTGCCATTTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5114 ATGCTGCCATTTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCG 5173

Qy         61 TTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACT 120
```

```
Db       5174 TTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACT 5233

Qy        121 ATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5234 ATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGAT 5293

Qy        181 AAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5294 AAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCA 5353

Qy        241 TTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5354 TTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTAC 5413

Qy        301 ATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5414 ATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGA 5473

Qy        361 GAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5474 GAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATT 5533

Qy        421 AATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5534 AATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGA 5593

Qy        481 TATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5594 TATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGA 5653

Qy        541 TTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5654 TTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGA 5713

Qy        601 GGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5714 GGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATG 5773

Qy        661 CTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5774 CTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAAT 5833

Qy        721 ATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5834 ATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCG 5893

Qy        781 CGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAGAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5894 CGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAGAA 5953

Qy        841 CATCCATCAGGTATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5954 CATCCATCAGGTATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGAT 6013

Qy        901 AATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6014 AATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGC 6073

Qy        961 TTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6074 TTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGA 6133

Qy       1021 ATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6134 ATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATA 6193

Qy       1081 TATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6194 TATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTAT 6253

Qy       1141 TGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGACGGTGCACGCGATAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6254 TGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGACGGTGCACGCGATAA 6313


RESULT 3
AX029263
LOCUS       AX029263               12446 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 3 from Patent WO9850531.
ACCESSION   AX029263
VERSION     AX029263.1  GI:10190114
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 3 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .12446
                     /organism="Salmonella enterica"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match           100.0%;  Score 1200;  DB 6;  Length 12446;
  Best Local Similarity  100.0%;  Pred. No. 7.1e-197;
  Matches 1200;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCTGCCATTTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5114 ATGCTGCCATTTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCG 5173

Qy         61 TTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5174 TTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACT 5233

Qy        121 ATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5234 ATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGAT 5293

Qy        181 AAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5294 AAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCA 5353

Qy        241 TTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5354 TTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTAC 5413

Qy        301 ATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5414 ATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGA 5473

Qy        361 GAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5474 GAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATT 5533

Qy        421 AATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5534 AATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGA 5593

Qy        481 TATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5594 TATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGA 5653

Qy        541 TTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5654 TTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGA 5713

Qy        601 GGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       5714 GGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATG 5773

Qy        661 CTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5774 CTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAAT 5833

Qy        721 ATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5834 ATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCG 5893

Qy        781 CGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAGAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5894 CGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAGAA 5953

Qy        841 CATCCATCAGGTATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5954 CATCCATCAGGTATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGAT 6013

Qy        901 AATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6014 AATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGC 6073

Qy        961 TTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6074 TTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGA 6133

Qy       1021 ATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6134 ATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATA 6193

Qy       1081 TATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6194 TATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTAT 6253

Qy       1141 TGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGACGGTGCACGCGATAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6254 TGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGACGGTGCACGCGATAA 6313
```

```
RESULT 4
BX890608/c
LOCUS       BX890608              194784 bp    DNA     linear   VRT 29-APR-2005
DEFINITION  Zebrafish DNA sequence from clone CH211-195J11 in linkage group 2,
            complete sequence.
ACCESSION   BX890608
VERSION     BX890608.14  GI:62989979
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 194784)
  AUTHORS   Lovell,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 30, 2005 this sequence version replaced gi:62751002.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml
            CH211-195J11 is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .194784
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-195J11"
                     /clone_lib="CHORI-211"
ORIGIN

  Query Match             5.7%;  Score 68.8;  DB 5;  Length 194784;
  Best Local Similarity   44.7%;  Pred. No. 0.0069;
  Matches  313;  Conservative    0;  Mismatches  382;  Indels    5;  Gaps    1;

Qy         36 TGTACTCAATGTATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAA 95
              |||  | | | || ||||| |     |  |  | |||    |  | ||   ||  |  ||
Db      83831 TGTGGTTATTATAGTTTTTATTATTATTATTGTTGTTATTGTCATTATCATTAATATTAA 83772

Qy         96 AACTTATTTCCCGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATGGGTAAT 155
              | |  | |     | | | ||| |||||||| |||||      || || |||   || |
Db      83771 TATTGTTGTTATCATTGCTATTATTATTTTTTTATGGTTATTATTATTATTATTATTATT 83712

Qy        156 AGCGTTAACTATATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAGTTATAT 215
              |   | |    |||| ||| |         |   |  |||  ||  |     |   |   |
Db      83711 ATTATCATTATTATTATCATCATTATCCTTAATATTAATATTGTTGTTATTGTTGTTCTT 83652

Qy        216 TTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTT 275
              |||||     || | ||  ||   ||  |  | |||  | ||| ||   |   ||| ||
Db      83651 ATTATTATTTTCATTATTTTTATTGTTGTTATTATTATTATTATTATCATCATTATCATT 83592

Qy        276 AAA-----TATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACA 330
              ||       |||||      ||  ||    |  || |||  | | ||  || | |  |||  |
Db      83591 AATATTATTATTGTTGTTATTATTGTTTTTATTATTAATGTCATTATTTTTATTGTTATT 83532

Qy        331 ATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAG 390
              ||   |||||||  |       || | |        |||| |   |  |       |  ||
Db      83531 ATTATTATTATTATTACTAATATTATTATTATTATCATCATCATCATTATTATTTTTGTT 83472

Qy        391 CATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACA 450
              |  |          |  | |  || |||    |    |  |   ||  ||     ||
Db      83471 TTTATTATTATTATTAACATTATTTTTATTGTCATTATTATTATTATCATCATTATTATT 83412

Qy        451 GGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGT 510
                | ||| |   |  | | | |       ||| |||    |   |  |  |      |
Db      83411 ATTATTATTATTATTATTATTATCATTATTATCATCATTATCCTTAATATTAATATT 83352

Qy        511 ATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTA 570
              | |    | ||   |      ||      |  |  |  || || ||| | || |  |
Db      83351 GTTGTTATTGTTGTTCTTATTATTATTTTCATTATTTTTATTGTTGTTATTATTATTATT 83292

Qy        571 TTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTAC 630
              ||||    ||   ||| | | ||| ||  |  |  ||  |||||  |    ||| |  |
```

```
Db      83291 ATTATCATCATTATCATTAATATTATTATTGTTGTTATTATTGTTTTTATTATTAATGTC 83232

Qy        631 GCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATA 690
              ||| |||  ||   | |   |  ||  || ||  |   || | | |  || | ||

Db      83231 ATTATTTTTATTGTTATTATTATTATTATTATTACTAATATTATTATTATTATCATC 83172

Qy        691 ATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACA 730
              || ||      || |      ||| || || |||| || |||

Db      83171 ATCATCATTATTATTTTTGTTTTTATTATTATTATTAACA 83132


RESULT 5
BX899180
LOCUS       BX899180              182150 bp    DNA     linear   VRT 06-AUG-2005
DEFINITION  Zebrafish DNA sequence from clone DKEY-287G19 in linkage group 21,
            complete sequence.
ACCESSION   BX899180
VERSION     BX899180.11  GI:71893330
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 182150)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Aug 6, 2005 this sequence version replaced gi:71833912.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep
            Clone-derived Zebrafish pUC subclones occasionally display
            inconsistency over the length of mononucleotide A/T runs and
            conserved TA repeats. Where this is found the longest good quality
            representation will be submitted.
            Any regions longer than 1kb tagged as misc-feature 'unsure' are
            part of a tandem repeat of more than 10kb in length where it has
            not been possible to anchor the base differences between repeat
            copies.  The region has been built up based on the repeat element
            to match the total size of repeat indicated by restriction digest,
            but repeat copies may not be in the correct order and the usual
            finishing criteria may not apply.
            DKEY-287G19 is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5.
FEATURES             Location/Qualifiers
     source          1. .182150
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="21"
                     /clone="DKEY-287G19"
                     /clone_lib="DanioKey"
ORIGIN
```

```
  Query Match             5.4%;  Score 65.2;  DB 5;  Length 182150;
  Best Local Similarity   46.2%;  Pred. No. 0.029;
  Matches  252;  Conservative    0;  Mismatches  293;  Indels    1;  Gaps    1;

Qy        215 TTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCT 274
              | |||||  |   | | ||  |    ||  |  | |||  | ||| | |
Db      17757 TGTTATTATCACTATTATTATTGATGTTGATATTATTATTATTATTACATATTATTATTA 17816

Qy        275 TAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATAT 334
              | | ||||   |  ||  ||      || ||| || | ||   | | | ||     ||
Db      17817 TTATTATTATTATTATTATTATTATTATTATTATTATTGATGTTGATGTTGATATTG 17876

Qy        335 CTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATT 394
              |||||||  | |  | ||  |||  ||  | ||||   |  |   |||| || |  |||
Db      17877 TTATTATTACTATTATTATTGATGTTGATATTATTATTATTATTACATATTATTATTATT 17936

Qy        395 ATGCAGGAGAACAAT-TAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGC 453
              ||       |  | | | |   | ||||||    ||| ||  || |||     |||
Db      17937 ATTGTTATTATTATTATTATTATTATTAATGTTGATGTTGATATTGTCATTATTACTATT 17996

Qy        454 GGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATG 513
              | ||  || || | | | | ||     |||  | |   |   | || |  |   |||
Db      17997 ATTATTGATGTTGATATTATTATTATTATTATTATTATTATTATTATTATTATTATT 18056

Qy        514 GAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTT 573
              | ||    | |  |||   |  |  | ||  | |||||| | | ||  |  ||
Db      18057 ATTATTATTATTGATGTTGATATTTGTATTATTATTATCATTATTAATATTATTATTATT 18116

Qy        574 ATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCA 633
              ||    ||    || | | ||  |   |      |  || ||  |    ||||| |
Db      18117 ATTATTATTATTATTATTATTGATGTTGATATTTTTATTATTATTATCATTATTAATATT 18176

Qy        634 CTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATT 693
              ||  ||| |||   | |   |  ||  ||  |  |   || | | |  | | || |||
Db      18177 ATTATTATTATTATTATTATTATTATTATTAATATTATTATCATTATTATCATTATTATT 18236

Qy        694 ATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGATCGGC 753
              ||    |||   |    |  || || |||| ||   ||  ||    | | | ||   |
Db      18237 ATTGATGTTGATGTTGATATTATTATTATTATTATTAATATTAATATTAATATTATTATC 18296

Qy        754 AGGTTT 759
              |  | |
Db      18297 ATTTAT 18302


RESULT 6
AR579680
LOCUS       AR579680               1141 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 22 from patent US 6784342.
ACCESSION   AR579680
VERSION     AR579680.1  GI:56583130
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1141)
  AUTHORS   Kunst,L. and Clemens,S.
  TITLE     Regulation of embryonic transcription in plants
  JOURNAL   Patent: US 6784342-A 22 31-AUG-2004;
            The University of British Columbia; Vancouver;
            CAX;
FEATURES             Location/Qualifiers
     source          1. .1141
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             5.3%;  Score 63.8;  DB 6;  Length 1141;
  Best Local Similarity   12.1%;  Pred. No. 0.31;
  Matches  110;  Conservative  335;  Mismatches  457;  Indels    6;  Gaps    3;

Qy        274 TTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATA 333
              ::::: ::: :::: :||      |:|: : ::: | :::::   ||| ::: : :::
```

```
Db        127 YWRWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNATGWDDDTKYHMWNN 186

Qy        334 TCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCAT 393

Db        187 NGCBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRKVRRWVRTRG 246

Qy        394 TATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGC 453

Db        247 RMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRAKYWGWNRABVNSTC 306

Qy        454 GGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATG 513

Db        307 TTWKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNNWTYKKARHBARW 366

Qy        514 GAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTT 573

Db        367 DWVWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYN 426

Qy        574 ATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCA 633

Db        427 NNNNGGRTYYG-WTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKTYWNNCYWK 485

Qy        634 CTT----TATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAAT 689

Db        486 SMTNGKSHRBAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBVCSKWWNNYAAWYT 545

Qy        690 AATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGAT 749

Db        546 KSSWNYTSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWSHTWBHBRA 605

Qy        750 CGGCAGGTTTACAGATTCGTCTCAGGCATCGCGTGGATCTACACAGGGGCGGATAGATAT 809

Db        606 GAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNNNNNNATCARDDYYAA 665

Qy        810 GGCAATTAATTCATTAAACTTCCTGTCAGAACATCCATCAGGTATAGGTCTGGGTACTCA 869

Db        666 SRWYAMANAKWYYYKBAANNAYYT-HANNWWGCWNNATDTRRTMWKNNNNNNAGTWKNNN 724

Qy        870 AGGTTCAGGAAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGATAATTATTTTTT 929

Db        725 NNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADAAABTTDKRNNGAYTKYTTTN 784

Qy        930 CTGGATCGCCCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCTGGCAAG 989

Db        785 NNNTYRGVVTNTAARDGWANNNNNNNNNNNNNNGWSDMWVTWWAYANYGTNNNNNNNNN 844

Qy        990 TCAATTTTATTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAGCAATATGCA 1049

Db        845 AYAWWTNKWYYTTDDRWRBAYTNNNNNNRMAYYGAYADDYAYYMSDTCDAWMKWDATKMN 904

Qy       1050 CTATAGATTATATTTTCTCTTTGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGC 1109

Db        905 NATTYNRGTAWRTNNNNNNMTMKTKYYBHAAWNNNNNNGKMCTAHTWWVCKATKTTKGCW 964

Qy       1110 ACCTTCGTCATCAACTTTTTCTATATATTATTGGACAGTTTTAGCTTTGATTCCATTTTT 1169

Db        965 MNCTTTCRKYKNNCTWYTWMTTTRTTWYAATRWKTNNATGSMTRCNATGWKNNNYWTGWK 1024

Qy       1170 AAAATTAA 1177

Db       1025 TRWTAYRM 1032


RESULT 7
AX083744
LOCUS       AX083744               1141 bp    DNA     linear   PAT 28-FEB-2001
DEFINITION  Sequence 22 from Patent WO0111061.
ACCESSION   AX083744
VERSION     AX083744.1  GI:13185472
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Kunst,L. and Clemens,S.
  TITLE     Regulation of embryonic transcription in plants
  JOURNAL   Patent: WO 0111061-A 22 15-FEB-2001;
            UNIVERSITY OF BRITISH COLUMBIA (CA)
FEATURES             Location/Qualifiers
     source          1. .1141
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
     promoter        1. .1141
                     /note="consensus sequence of A.t., L.a., and B.n. FAE1
                     promoters"
ORIGIN

  Query Match             5.3%;  Score 63.8;  DB 6;  Length 1141;
  Best Local Similarity   12.1%;  Pred. No. 0.31;
  Matches  110;  Conservative  335;  Mismatches  457;  Indels    6;  Gaps    3;

Qy        274 TTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATA 333

Db        127 YWRWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNATGWDDDTKYHMWNN 186

Qy        334 TCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCAT 393

Db        187 NGCBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRKVRRWVRTRG 246

Qy        394 TATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGC 453

Db        247 RMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRAKYWGWNRABVNSTC 306

Qy        454 GGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATG 513

Db        307 TTWKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNNWTYKKARHBARW 366

Qy        514 GAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTT 573

Db        367 DWVWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYN 426

Qy        574 ATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCA 633

Db        427 NNNNGGRTYYG-WTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKTYWNNCYWK 485

Qy        634 CTT----TATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAAT 689

Db        486 SMTNGKSHRBAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBVCSKWWNNYAAWYT 545

Qy        690 AATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGAT 749

Db        546 KSSWNYTSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWSHTWBHBRA 605

Qy        750 CGGCAGGTTTACAGATTCGTCTCAGGCATCGCGTGGATCTACACAGGGGCGGATAGATAT 809

Db        606 GAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNNNNNNATCARDDYYAA 665

Qy        810 GGCAATTAATTCATTAAACTTCCTGTCAGAACATCCATCAGGTATAGGTCTGGGTACTCA 869

Db        666 SRWYAMANAKWYYYKBAANNAYYT-HANNWWGCWNNATDTRRTMWKNNNNNNAGTWKNNN 724

Qy        870 AGGTTCAGGAAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGATAATTATTTTTT 929

Db        725 NNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADAAABTTDKRNNGAYTKYTTTN 784

Qy        930 CTGGATCGCCCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCTGGCAAG 989

Db        785 NNNTYRGVVTNTAARDGWANNNNNNNNNNNNNNGWSDMWVTWWAYANYGTNNNNNNNNN 844

Qy        990 TCAATTTTATTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAGCAATATGCA 1049

Db        845 AYAWWTNKWYYTTDDRWRBAYTNNNNNNRMAYYGAYADDYAYYMSDTCDAWMKWDATKMN 904

Qy       1050 CTATAGATTATATTTTCTCTTTGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGC 1109
```

```
Db        905 NATTYNRGTAWRTNNNNNNMTMKTKYYBHAAWNNNNNNGKMCTAHTWWVCKATKTTKGCW 964

Qy       1110 ACCTTCGTCATCAACTTTTTCTATATATTATTGGACAGTTTTAGCTTTGATTCCATTTTT 1169
              : |||      :    ::|:  |   |:: ||  :    |    :|  ||::    :| ::
Db        965 MNCTTTCRKYKNNCTWYTWMTTTRTTWYAATRWKTNNATGSMTRCNATGWKNNNYWTGWK 1024

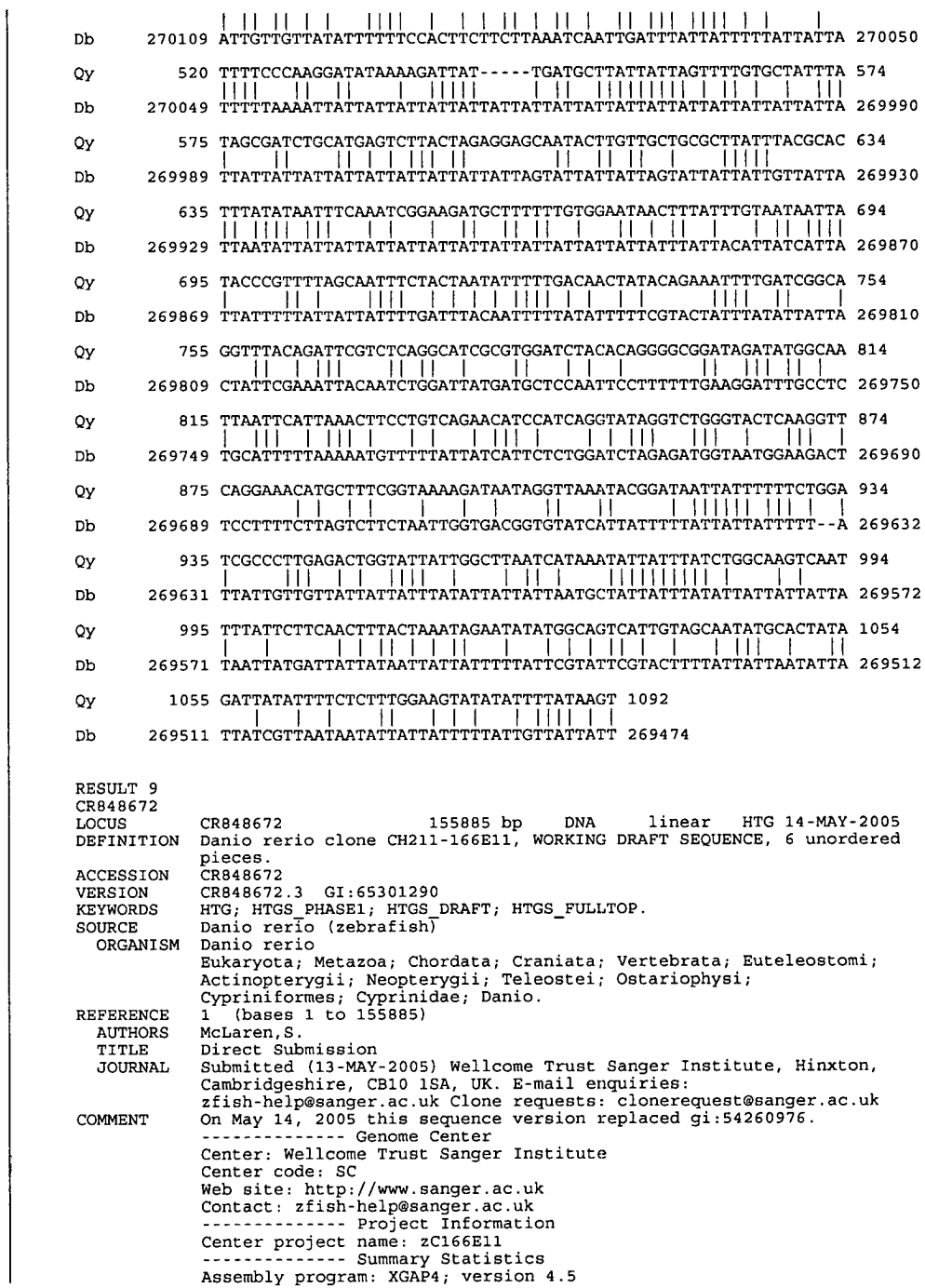
Qy       1170 AAAATTAA 1177
                ::  :::
Db       1025 TRWTAYRM 1032


RESULT 8
AC116979/c
LOCUS       AC116979              331039 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 6445720-6776760 strain
            AX4, complete sequence.
ACCESSION   AC116979
VERSION     AC116979.3  GI:42733765
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 331039)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 331039)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 331039)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 331039)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 331039)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Feb 21, 2004 this sequence version replaced gi:28829703.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .331039
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="6445720-6776760"
     CDS             complement(join(912. .1165,1280. .1427))
                     /note="GeneID exon scores (in order of location ranges):
                     30.46, 12.19 - GSCJ_ID dd_02783"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38697.1"
                     /db_xref="GI:42733766"
                     /translation="MSEKTFTRQEVAKHCSLNSLWIIYNDDVFDVTNFVVEHPGGEEV
                     LKGNGGKDATQEFDDVGHSASAIAKMQSLRIGRIEGAKPREEKKKEIKKTTTTTSAPK
                     QQESAGLGLLKIPLIIIVLAIAAYFFMGDQQ"
     CDS             complement(join(2345. .3098,3189. .3526))
                     /note="GeneID exon scores (in order of location ranges):
                     98.40, 31.82 - GSCJ_ID dd_02784"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Protein disulfide isomerase precursor"
                     /protein_id="AAO52220.2"
                     /db_xref="GI:42733767"
                     /translation="MKILLFVTLIALAFVALCSAEGNVVVLSPDNFDTVVDGSKTVFV
                     KFYAPWCGHCKKLAPDFEILADTFAPVSNKVVIAKVDCDQADNKALCSKYDVSGYPTL
                     KIFDKSTTAKDYNGARSVDELLTYINNHAKTNVKVKKAPSNVVDLSPSNFDSVVLDKS
                     KNVLVEFYAPWCGHCKKLMPDYEILGNTYANEKDVVIAKIDCDAADNKAICSKYGVTG
                     FPTLKWFGKQSKDGEKYEQGRDLDTFINYINKQAGVNRVKGGKLAVGAGRVEQLDTIA
                     TEFIAAAAEVRKELVKKAQTVVDSLPEELRTEGSYYVKVMKTIAEKSIDFVTTEIARI
                     TKLVSGSMSGKKADEFAKKLNILESFKSK"
     CDS             join(5795. .6179,6292. .6567,6659. .6881,7332. .7363,
                     7523. .8215,8284. .10451)
                     /note="GeneID exon scores (in order of location ranges):
                     14.11, 5.65, 10.80, 0.73, 45.28, 141.12 - GSCJ_ID
                     dd_02785"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Hypothetical 127.0 kDa protein"
                     /protein_id="AAO52221.2"
                     /db_xref="GI:42733768"
                     /translation="MNINLLKFFIIYYFCFVLINFCNCQTLKITDISNPKNYKVYPQY
                     SDREFCDPSFEILCYVEPAMDEGDQLMFTTDFDYSEMIRSDRTGSIIGITHTALSFGS
                     YTSNIVCYLLNNPTIKVDLNISYSCEMIDYELMKVKILKPNYFKKNNLYGNGIIQLIG
                     LKYPLNNLVIFGGSAYPLNPNYFLISLEELDGGTVENYNISLLFYGPHTIVVTIPAST
                     YLDYNLDISNSKINYYPNETTNVSEFGLGCSTVYSVTINKTSNDYEPYLYVDDGNSKN
                     ADFTPIYQDENFITYLGQFNSFVTVTSPYAQILNSSILSTLLEFECKLLFDESYLFRF
                     KVNTNGYGASIYLGSDITLRYESVVDSESNSSVLEFETVVDLYKDSITKFTISNLLSN
                     RINYLRGEYYSISPLLKFEIPFGELNPVLITNISFLYNNIDVTNKSVNNIIYFSYEGE
                     FDKGSSFALLLNDPVSILKNGDLDSSDFRFSKWNSSTSMFQIEFTVPANTQPGHIPFK
                     LYGNSNPNYNPIDSTMLPLSAQLFIISSNFDSYGPIFNKVEKINTTNEFGWKFVIGDK
                     INGFDYGDIIVRGEMDSSTYKFHLTPLNLTSGDKYNGEYQINIKISILCASQNYIITD
                     VYLFDTRGNFAKFSVSEQCESIRNPFINFLGDSTINKIYKKCTTENDGYDISPPELTL
                     FSAVKTILNDSQTIVIAFDFQAVDLDTGLKNEQYPIVYIESTELQIIQCESKIKSINS
                     KTANYTCEIELPVGFGYNSDIVFSVYGFINNGGYYSGFPSTYISSAFPIYSMSNIQLV
                     RKIEILRTSQISSTGGELWIVGRGFKTSIQYVQIKYSTDLVFSQISFPSIIFSTALLI
                     LDIDPTDKPFTIRVVTKNGQQSNEFIVTPLFSFLSVSPLNTPTPTTTTQLPTPTNKPQ
                     TCLGEPICGGPKQGFCSSKGCVCYQPWIGNDCNAQVVIIPQPSTNTSQPSTEIPIIDN
                     NNNQQPSGINYLYKSLISIVSIRELDFNGNQINLYPFDRWIFTQINQNKSQYFTSIKN
                     NSTTPFNTNITVILEWFKDTSTIEFANSSITMYPSSIKYTIEITEYKFSNQLNNLQLV
                     MSASFESSKTKEICSLKDFGETSNGDNSNYFKIQINDHSLYGRFIKRAIIDSKISTIE
                     NVLLDSKMNSIQTASSSQSYIGITIPNYQKSIIIDPDFSVLIDSKSTSKNDENSVCTP
                     NQSKLTTPQLIGIIIGSIAFAAVIAISIAYHIIKKKKDTKFKNGLQNKLKTFKS"
     CDS             join(11030. .11042,11114. .11283)
                     /note="GeneID exon scores (in order of location ranges):
                     -1.76, 12.26 - GSCJ_ID dd_02786"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52222.2"
                     /db_xref="GI:42733769"
                     /translation="MTILKSISSMGISNRKNSFSFVSTNSMIETNQNVNSVSKSGYSK
                     LIKGAALMAEGISGFF"
     CDS             complement(11515. .12534)
                     /note="GeneID exon scores (in order of location ranges):
                     87.29 - GSCJ_ID dd_02787"
                     /codon_start=1
                     /product="similar to Burkholderia sp. DBT1. Putative
                     oxidoreductase"
                     /protein_id="AAM34303.3"
                     /db_xref="GI:42733770"
                     /translation="MLPITNKVALITKKGPSESFSIVDNEIPKPKGNEVLIKVDYVGL
                     NRCDLYFSLGLYLPFENYPNVSGFEASGTVVEFANDELENSSKFKVGDKVSVIYTCDH
                     RKYGTFAEYAVFPIGCLVKNNELSSQKEIAANWISFFTAYYGLVEFAQIKKGDYVIIT
                     AGSAGTSLAGIAIAKFFGAHTIVTSRNEKKKQQILDYGADYFVSLDTEDIVKRTFEIT
```

```
                NGKGANIIYDSVGGQSAQKLFESSSQFGKIVFYGNMNHSEPTPLPILIGLKNYLTIKV
                LNTNEYLYQPEFNKKAIEFLNSNFHHFKSIVGKEFIGIDSLVDALKYLETGDLFGKVV
                VKVNN"
     CDS        complement(join(13231. .14884,15364. .15368))
                /note="GeneID exon scores (in order of location ranges):
                104.49, -7.13 - GSCJ_ID dd_02788"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAM34304.3"
                /db_xref="GI:42733771"
                /translation="MINPNLLKPLFIENFCSSSQSWLDLIKNRKLNFYTITCLTHYYS
                TVKVLYFDIKSIFENDKKQNENQQNEKQQQQQQQQQQQQQNEKSLNDRKSIFKKSYH
                IIRLLIEMNDIINYKDIRIYLPNQSEERKLLMKIREGEEIDIQSCFELVQSKYLEFES
                KFNELKKNSTNDSLIKLNSARALHVLNDWVVDIRRNSIENSNNLNSFKKYKFENENHH
                DESNNSFKIAMDLMNKNNIKGTLIYFGPSGSHLHGLQDDDHNDENSNDNEKIDWIGIY
                AAPTDDIVSLFPLALKHDSSGFIMQSNEHIGINLNPAPDRPNNQNKKDSFSKGIILYE
                LKYALQLIVNCSYRVLECVFMSSNQVLNNHPFKNKFYITDAWKDFINQVLINDNNNNK
                TFSKLSNSNLIQQYWGLAKGISMKNSTIFKMVKSNLLRKKDDAITIAEEDLVTDDSIS
                KNLFFLNYLLIQCNTILNGIVFNENNTIINSPNNEILLKEKEIPSIFIDKQSKDYQFL
                MSLKHNKLNKLSIIQSLIIDSDSILEEIRIKLDTLKKIQRIDLV"
     CDS        join(17515. .17664,18062. .19243)
                /note="GeneID exon scores (in order of location ranges):
                1.18, 93.30 - GSCJ_ID dd_00463"
                /codon_start=1
                /product="similar to mitochondrial carrier protein family;
                protein id: At1g14560.1 [Arabidopsis thaliana]"
                /protein_id="AAO52223.2"
                /db_xref="GI:42733772"
                /translation="MNSSDFKKSFKESTENNSNTYRPSKTLNTYKSLKNGNNIANSLI
                IQLNNNYLVSGSIAGAISRSATAGFERLTIIQQVQGMSQNLSQGYVGCIAAMKEMVKR
                EGFKSIWKGNGANIVKVSPNSGIRFLTYEFCKKHFLDNSSNHPSSSSIENGIDGNGVG
                CGSGSEMKMTVPQTMFSGAMAGLTSTFFTYPLDVVRIRLSLQGSCSNDYAAHRYNGIT
                HSFFKIHKDEGVKGLYKGLGTSIASIVPWVSISFATYEGFKIICKKMILNYQISSSSL
                STTTTTPSSINNNNNNNNKNNNSFIYENELGENGINLTNTSGCSTMASTMPSSLLINS
                VASDENELKKGVNMICDFVCGALSGAVTMTVCYPLDVLRRRMMIQGIGGNKVLYKNGW
                DATKKILSNEGLVAFYHGIIPAYFKVVPTVAISFAVYEICKDLGSNKYQQK"
     CDS        join(20299. .20406,20551. .20660,20960. .21569,21708. .21828,
                22984. .24188)
                /note="GeneID exon scores (in order of location ranges):
                9.06, -2.15, 55.50, 6.70, 58.02 - GSCJ_ID dd_01980"
                /pseudo
                /codon_start=1
     CDS        join(24256. .24266,24388. .24538,24623. .24750,24902. .27884)
                /note="GeneID exon scores (in order of location ranges):
                -2.62, 6.90, 11.20, 259.32 - GSCJ_ID dd_01978"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). Histidine kinase DhkF"
                /protein_id="AAO52224.2"
                /db_xref="GI:42733773"
                /translation="MILKFIIQQCVFKVAKIFSDNLIQIFMDISSVYSSSTLFFSKLS
                LLMCKVFNSKCAGISVLNKDGLFDTIAFCSSTQELENFSYSLKKISKNITYYDYMEIP
                LKDSKEFYFGDIVLFGDQTLNPNILDTNILTFITNRTSMEFEKKKVSDQLIIAKNLLD
                QSPSCSCLVSRNGKIIRKIGAFYERLLGLEIGENISSVEKSSDISTILKQVCFENNNL
                STETTITKKNGEKYPAEVFVKEISDIHSNSIGIMIIVRDITDQIRLKEMNIELQKKSE
                LEQKRNQELMEARDLALTATKIKSQFLATISHEIRTPLNGIITMGEMLLSTSPLNTEQ
                HDIAETIFGSSELLLSITSDILDFSKIEASKLELEMIEFDFIGCLEGIGKTIGVSITN
                KPIEIAFLMDTDIPHRLIGDPNRLIQIMLNMGTNAVKYTDRGHIVFRISVISREQNRC
                KIKISIDDSGIGIGEDQRAHLFEPFHQIDSSSTRKYGGSGLGLAISSKLAKLMGGEVI
                LERSKPGVGSLFSVTLNFEQIGSNTLKSLLPDKSFSRNKTCIILDNYEYTASIASQRF
                DQIFEESNIILVKESTVKKFFELVQEYKINGDVNKIDLKNLDPELLPFFDSSLICIIV
                FHRFMDNLDIFLKYVKDFIEFYKKKIVVALGINHKNFKNLPKRRDFLIFKKPISSTNL
                IKVSNMARKIIPKSSSSSNLIQTISQIDNQQQQQQQQLQQQEQEQQHQQQQLQQEQQF
                VTISPHSDSSEKKTTPKKDRGKYDFNISPLRIDRFGTESTSSPKIKLYSDTSSDSGES
                DEFEFSENLRELKSDGELVVLKNQPYKEQRLSNIQLINDPIVTPPTPKANSEVVSTKA
                ESTFFQSSSITIKETRSDSVNSVNSVNSVNSVNSLNSENSNNSLDVGIRTRPRILLVD
                DNAVNRKVVKLQLKKLGYDCDTATNGFEGFEMQKKDNYELIFMDLNMPLCDGSMASKL

  Query Match             5.3%;  Score 63.2;  DB 2;  Length 331039;
  Best Local Similarity   45.9%;  Pred. No. 0.052;
  Matches  293;  Conservative    0;  Mismatches  338;  Indels    7;  Gaps    2;

Qy        460 AGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGT 519
```



```
Db     270109 ATTGTTGTTATATTTTTTCCACTTCTTCTTAAATCAATTGATTTATTATTTTTATTATTA 270050

Qy        520 TTTTCCCAAGGATATAAAAGATTAT-----TGATGCTTATTATTAGTTTTGTGCTATTTA 574

Db     270049 TTTTTAAAATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTA 269990

Qy        575 TAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCAC 634

Db     269989 TTATTATTATTATTATTATTATTATTATTAGTATTATTATTAGTATTATTATTGTTATTA 269930

Qy        635 TTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTA 694

Db     269929 TTAATATTATTATTATTATTATTATTATTATTATTATTATTTATTACATATCATTA 269870

Qy        695 TACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGATCGGCA 754

Db     269869 TTATTTTTATTATTATTTTGATTTACAATTTTTATATTTTTCGTACTATTTATATTATTA 269810

Qy        755 GGTTTACAGATTCGTCTCAGGCATCGCGTGGATCTACACAGGGGCGGATAGATATGGCAA 814

Db     269809 CTATTCGAAATTACAATCTGGATTATGATGCTCCAATTCCTTTTTTGAAGGATTTGCCTC 269750

Qy        815 TTAATTCATTAAACTTCCTGTCAGAACATCCATCAGGTATAGGTCTGGGTACTCAAGGTT 874

Db     269749 TGCATTTTTAAAAATGTTTTTATTATCATTCTCTGGATCTAGAGATGGTAATGGAAGACT 269690

Qy        875 CAGGAAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGATAATTATTTTTTCTGGA 934

Db     269689 TCCTTTTCTTAGTCTTCTAATTGGTGACGGTGTATCATTATTTTTATTATTATTTTT--A 269632

Qy        935 TCGCCCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCTGGCAAGTCAAT 994

Db     269631 TTATTGTTGTTATTATTATTTATATTATTATTAATGCTATTATTTATATTATTATTATTA 269572

Qy        995 TTTATTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAGCAATATGCACTATA 1054

Db     269571 TAATTATGATTATTATAATTATTATTTTTATTCGTATTCGTACTTTTATTATTAATATTA 269512

Qy       1055 GATTATATTTTCTCTTTGGAAGTATATATTTTATAAGT 1092

Db     269511 TTATCGTTAATAATATTATTATTTTTATTGTTATTATT 269474
```

```
RESULT 9
CR848672
LOCUS       CR848672              155885 bp    DNA     linear   HTG 14-MAY-2005
DEFINITION  Danio rerio clone CH211-166E11, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   CR848672
VERSION     CR848672.3  GI:65301290
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 155885)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 14, 2005 this sequence version replaced gi:54260976.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC166E11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
```

```
           Chemistry: Dye-terminator; 100% of reads
           Consensus quality: 153972 bases at least Q40
           Consensus quality: 154337 bases at least Q30
           Consensus quality: 154522 bases at least Q20
           Insert size: 155385; sum-of-contigs
           Insert size: 153983; 2.8% error; agarose-fp
           Quality coverage: 7.38x in Q20 bases; sum-of-contigs Quality
           coverage: 7.70x in Q20 bases; agarose-fp
           --------------.
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 6 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *        1    10134: contig of 10134 bp in length
           *    10135    10234: gap of 100 bp
           *    10235   111588: contig of 101354 bp in length
           *   111589   111688: gap of 100 bp
           *   111689   123934: contig of 12246 bp in length
           *   123935   124034: gap of 100 bp
           *   124035   133984: contig of 9950 bp in length
           *   133985   134084: gap of 100 bp
           *   134085   149579: contig of 15495 bp in length
           *   149580   149679: gap of 100 bp
           *   149680   155885: contig of 6206 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155885
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-166E11"
                     /clone_lib="CHORI-211"
     misc_feature    1. .10134
                     /note="assembly_fragment:00183
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    10235. .111588
                     /note="assembly_fragment:00620
                     fragment_chain:1"
     misc_feature    111689. .123934
                     /note="assembly_fragment:00082
                     fragment_chain:1"
     misc_feature    124035. .133984
                     /note="assembly_fragment:00296
                     fragment_chain:1"
     misc_feature    134085. .149579
                     /note="assembly_fragment:00428
                     fragment_chain:1"
     misc_feature    149680. .155885
                     /note="assembly_fragment:00029
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             5.2%;  Score 62.8;  DB 14;  Length 155885;
  Best Local Similarity   47.7%;  Pred. No. 0.08;
  Matches  249;  Conservative    0;  Mismatches  267;  Indels    6;  Gaps    2;

Qy        228 AGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAA 287
              |  || |||  |||| |  | |||| | ||||||       ||| || ||||||   |
Db     116173 ATTATTAATATTATTATTTGTATTAATATTATCATTATAATTATCATT-AATATTATTAT 116231

Qy        288 AATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGT 347
               ||  ||      || ||| || ||||  ||||  || | ||| ||  | |||||| | |
Db     116232 TATTATTATTATTATTATTGTTATAATTATTATAATTACTACAGTAATTTTTATTGCTAT 116291

Qy        348 ACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACA 407
               | || | |       | |  |   |  |     ||  | |   |  |  |     |

Db     116292 TATTATTATTATTATTGTAACAATTATTATTATTAATATTAATATTGTTTTAATGATAAT 116351

Qy        408 ATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGC 467
              | |||   ||||  |  |   ||      |   |      ||  |  | | ||  || ||
Db     116352 AATAATAATAATAATAAATAAT-----AAATATTATTATTATTGTTGTTGTTGTTGTT 116406

Qy        468 TCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCA 527
              | |   | ||      |||  | | |  |  | || | |   |||       | ||
Db     116407 TGTTGTTATTATTATTATTATTATAGCAATTATTATTATTATTATTATTATTGTTGTTGT 116466

Qy        528 AGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCAT 587
              |  ||  |   |  | ||    |||||| | || |  |  ||||||| ||     ||
Db     116467 TGTTATTATTATTATTATTATAGCAATTATTATTATTATTATTATTATAGCAATTATTAT 116526

Qy        588 GAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTC 647
              | | ||  |   |  |   |  || ||  |    |||||   ||| ||  ||| |||
Db     116527 TATTATTGTTGTTGTTGTTATTATTATTATTATTATTATAGCAATTATTATTATTAT 116586

Qy        648 AAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGC 707
              | |   |      || ||  |   || | | ||     | || |||||    ||| | |
Db     116587 TATTATTATAGCAATTATTATTATTATTATTATAGCAATTATTATTATTATTGTTGTTGT 116646

Qy        708 AATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGAT 749
                ||  || || |||| ||  |    || || |||| | ||
Db     116647 TGTTGTTATTATTATTATTATTATTATTATAGCAATTATTAT 116688


RESULT 10
BX928756/c
LOCUS       BX928756              157210 bp    DNA     linear   VRT 27-JUN-2005
DEFINITION  Zebrafish DNA sequence from clone CH211-113K9 in linkage group 21,
            complete sequence.
ACCESSION   BX928756
VERSION     BX928756.13  GI:67763634
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 157210)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Jun 14, 2005 this sequence version replaced gi:62868301.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Any regions longer than 1kb tagged as misc-feature 'unsure' are
            part of a tandem repeat of more than 10kb in length where it has
            not been possible to anchor the base differences between repeat
```

```
            copies.  The region has been built up based on the repeat element
            to match the total size of repeat indicated by restriction digest,
            but repeat copies may not be in the correct order and the usual
            finishing criteria may not apply. CH211-113K9 is from a CHORI-211
            BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .157210
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="21"
                     /clone="CH211-113K9"
                     /clone_lib="CHORI-211"
ORIGIN

  Query Match             5.1%;  Score 61.4;  DB 5;  Length 157210;
  Best Local Similarity   45.6%;  Pred. No. 0.14;
  Matches  326;  Conservative    0;  Mismatches  386;  Indels    3;  Gaps    3;

Qy         47 TATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCC 106
Db     134508 TATTATTACTATTATTATTATTATTTTTACTATTATTATTATTATTACTATTATTT 134449

Qy        107 CGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTA 166
Db     134448 TTTTACTTTTATTATTATTATTATTATTATCATTATTACTATTATTATTACTATTACTAT 134389

Qy        167 TATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGG 226
Db     134388 TATTTTATATTATTATTATTGCTATTACTATTACTTTATTATTATTATTATTATTATTAT 134329

Qy        227 CAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAA 286
Db     134328 TATTATTAATTTTATTATTAATTTTATTATTATTTTATATTTTTATTATTACTATTACAA 134269

Qy        287 AAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATG 346
Db     134268 TATTATTATTATTATTATTATTATTATTACTATTATTATTATTATTGTTATTGTTTTTAT 134209

Qy        347 TACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAAC 406
Db     134208 TATTTTTACTGTTATTATTATTATTATTACTATTATTTTTTTACTATTATATTGTTATTA 134149

Qy        407 AATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATG 466
Db     134148 TTATTACTATTATTATTTTTACTATTATTTTTTACTATTATTATTATTATCATCATCATT 134089

Qy        467 CTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCC 526
Db     134088 ATTATTATTATTATTAT-TACTATTATTATTACTATTACTAT-TATTTTATATTATTATT 134031

Qy        527 AAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCA 586
Db     134030 ATTGCTATTACTATTAATTTATTATTATTATTTTTTTTACTATTATTATTATTTTTT-TG 133972

Qy        587 TGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTT 646
Db     133971 TTATTATTATTATTATTACTATTATTATTACTATTATTATTTTTACTATTATTATTATTA 133912

Qy        647 CAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAG 706
Db     133911 TTATTATTAATATTATTATTATTATTATTATTATTATTATTATTAATATTATTATTTTTA 133852

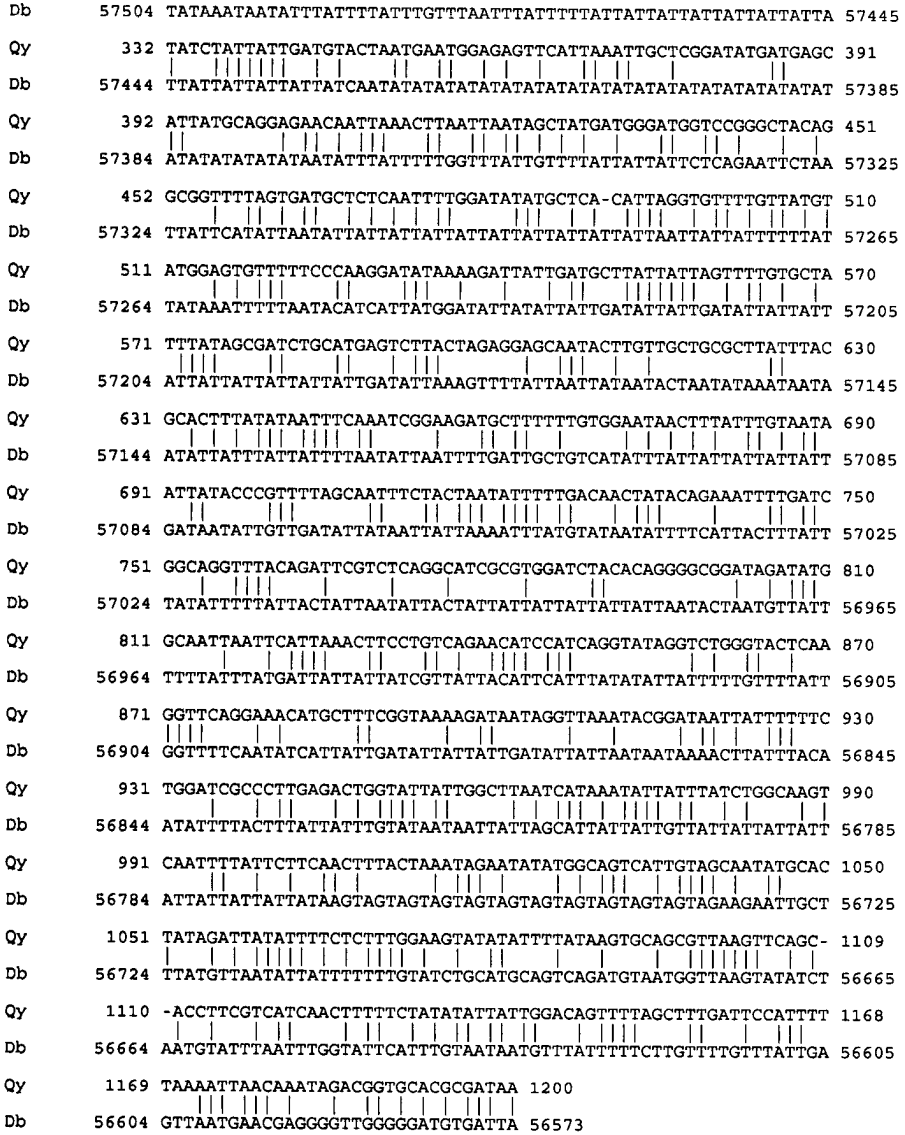
Qy        707 CAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTAC 761
Db     133851 CTATTATTATTATTATTATTATTATTACTATTATTATTATTATTATTTTTAC 133797


RESULT 11
CR388190/c
LOCUS       CR388190              142687 bp    DNA     linear   HTG 14-JUL-2005
DEFINITION  Danio rerio clone CH211-284E3, WORKING DRAFT SEQUENCE, 2 unordered
            pieces.
ACCESSION   CR388190
VERSION     CR388190.18  GI:70907945
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 142687)
  AUTHORS   Lovell,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Jul 15, 2005 this sequence version replaced gi:68162062.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC284E3
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 142483 bases at least Q40
            Consensus quality: 142539 bases at least Q30
            Consensus quality: 142547 bases at least Q20
            Insert size: 142587; sum-of-contigs
            Insert size: 142545; 13.5% error; agarose-fp
            Quality coverage: 11.05x in Q20 bases; sum-of-contigs Quality
            coverage: 13.04x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    54739: contig of 54739 bp in length
            *   54740    54839: gap of 100 bp
            *   54840   142687: contig of 87848 bp in length.
FEATURES             Location/Qualifiers
     source          1. .142687
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-284E3"
                     /clone_lib="CHORI-211"
     misc_feature    1. .54739
                     /note="assembly_fragment:00400
                     clone_end:SP6
                     vector_side:left"
     misc_feature    54840. .142687
                     /note="assembly_fragment:04252
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             5.0%;  Score 60.4;  DB 14;  Length 142687;
  Best Local Similarity   42.6%;  Pred. No. 0.21;
  Matches  423;  Conservative    0;  Mismatches  566;  Indels    3;  Gaps    2;

Qy        212 ATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAAT 271
Db      57564 ATATGTTTTACGCAGCAGATGCCCTTCCAGCTGCAACCCATCACTGGGAATTGTTGATAT 57505

Qy        272 CCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAA 331
```



RESULT 12
CR847804/c
LOCUS CR847804 202836 bp DNA linear HTG 06-AUG-2005
DEFINITION Danio rerio chromosome 13 clone DKEY-85C24, WORKING DRAFT SEQUENCE, 6 unordered pieces.
ACCESSION CR847804
VERSION CR847804.22 GI:71893353
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 202836)
AUTHORS Donaldson,S.
TITLE Direct Submission
JOURNAL Submitted (05-AUG-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT On Aug 6, 2005 this sequence version replaced gi:68162468.
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
---------- Project Information
Center project name: zK85C24
---------- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 196537 bases at least Q40
Consensus quality: 196747 bases at least Q30
Consensus quality: 196867 bases at least Q20
Insert size: 202336; sum-of-contigs
Insert size: 169122; 21.7% error; agarose-fp
Quality coverage: 8.60x in Q20 bases; sum-of-contigs Quality coverage: 10.36x in Q20 bases; agarose-fp
---------- .
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the the accession number will
* be preserved.
* 1 127064: contig of 127064 bp in length
* 127065 127164: gap of 100 bp
* 127165 129409: contig of 2245 bp in length
* 129410 129509: gap of 100 bp
* 129510 135419: contig of 5910 bp in length
* 135420 135519: gap of 100 bp
* 135520 138213: contig of 2694 bp in length
* 138214 138313: gap of 100 bp
* 138314 141007: contig of 2694 bp in length
* 141008 141107: gap of 100 bp
* 141108 202836: contig of 61729 bp in length.

FEATURES Location/Qualifiers
source 1. .202836
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/chromosome="13"
/clone="DKEY-85C24"
/clone_lib="DanioKey"
misc_feature 1. .127064
/note="assembly_fragment:02797"
misc_feature 127165. .129409
/note="assembly_fragment:02899"
misc_feature 129510. .135419
/note="assembly_fragment:02940"
misc_feature 135520. .138213
/note="assembly_fragment:03023"
misc_feature 138314. .141007
/note="assembly_fragment:03024"
misc_feature 141108. .202836
/note="assembly_fragment:03033"
ORIGIN

Query Match 5.0%; Score 60.2; DB 14; Length 202836;
Best Local Similarity 45.0%; Pred. No. 0.2;
Matches 308; Conservative 0; Mismatches 373; Indels 4; Gaps 2;

```
Qy          41 TCAATGTATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTT 100
               | | | |  || |||      | |||  |  || |  ||   | |  ||      |  | |
Db       96426 TTATTATTATTGTTGCTATTATTGTTATCATTAGTATTAGTAGTAGTATTGTTATTATTA 96367

Qy         101 ATTTCCCGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGT 160
                |||    |    || |  |  || ||  |  |     || ||  |      ||  | | |
Db       96366 TTTTTTATTATTGTTGTTGTTGTTGTTATTATATTGTTATTATTATTATATCATCAT 96307

Qy         161 TAACTATATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTAT 220
               || || |||| | |       || |   |  | ||                  |   | || |
Db       96306 TATCTTTATTATTATTATTATAATTATTATTATTTTCATCATTATTATCATTATTATTGT 96247

Qy         221 TCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATA 280
               | || | | | ||  || ||||| | || |||  |||||  |         |  || | ||
Db       96246 TACCTGTATTATTATTATCATTATTAATGTTAGTAGTATTGTTGTTATTTTTATT-ATTA 96188

Qy         281 TTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTA 340
               ||   |  ||  ||      ||  || || | ||  ||||  ||   || ||   |||||
Db       96187 TTATTATTATTATTATTATTATTGTTGTTGTTATTATTATTATTATTACTATTATTATTA 96128

Qy         341 TTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAG 400
               ||  | |  | |   || |   | |||   | |      || ||    |  | ||
Db       96127 TTATTATTGTTGTTGTTGTTGTTGTTATTGTTATTATTATTGTTGTTATTATTACCA 96068

Qy         401 GAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTA 460
                  | | | |||| || | |  |  ||   | ||       ||       | | |
Db       96067 TTATTATTATTATCATTATTATTATCATTATTGTTTTTGTTATTAGTATTATTATTATCA 96008

Qy         461 GTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTT 520
                  || | | | | ||     |||  | |   | | |||| |  |   |||    | ||
Db       96007 TCATTGTTATTATTATTCTTATTATCATTATTATTGTTGTTATTATTATTATTAGTTATT 95948

Qy         521 TTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGA 580
               ||   |   ||     |  |||||| ||  |||| |||| | || |  | |||   |  |
Db       95947 ATTATTATTTATTATTATTATTATTTATTATTATCATTATTATTATTGTTTTT---GTTA 95891

Qy         581 TCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATA 640
               |  | || | | |||  |          || ||| ||       |||| | |   ||  ||
Db       95890 TTAGTATTATTATTATAATCATTATTATTCTTATTATCATTATTATTAATGTTTTTATTA 95831

Qy         641 TAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCG 700
               | |||     |   |  ||  || ||  |   |  | | |  || | || |||||
Db       95830 TTATTGTTGTTATTATAATTATTGTTACTATTAATATTATTATTATTATTATTATTATTA 95771

Qy         701 TTTTAGCAATTTCTACTAATATTTT 725
               || |    |||| | | ||| | ||
Db       95770 TTATTATTATTTATTCCAATTTATT 95746
```

RESULT 13
CR759891/c
LOCUS CR759891 202208 bp DNA linear HTG 10-JUN-2005
DEFINITION Danio rerio chromosome 24 clone DKEY-261I18, WORKING DRAFT SEQUENCE, 4 unordered pieces.
ACCESSION CR759891
VERSION CR759891.4 GI:67509007
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 202208)
AUTHORS McLaren,S.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT On Jun 10, 2005 this sequence version replaced gi:52000376.
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
---------- Project Information
Center project name: zK261I18
---------- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 200681 bases at least Q40
Consensus quality: 201061 bases at least Q30
Consensus quality: 201450 bases at least Q20
Insert size: 201908; sum-of-contigs
Insert size: 207763; 0.9% error; agarose-fp
Quality coverage: 6.20x in Q20 bases; sum-of-contigs Quality coverage: 6.02x in Q20 bases; agarose-fp
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 159147: contig of 159147 bp in length
* 159148 159247: gap of 100 bp
* 159248 163227: contig of 3980 bp in length
* 163228 163327: gap of 100 bp
* 163328 168724: contig of 5397 bp in length
* 168725 168824: gap of 100 bp
* 168825 202208: contig of 33384 bp in length.

FEATURES Location/Qualifiers
source 1. .202208
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/chromosome="24"
/clone="DKEY-261I18"
/clone_lib="DanioKey"
misc_feature 1. .159147
/note="assembly_fragment:00385
fragment_chain:1"
misc_feature 159248. .163227
/note="assembly_fragment:00005
fragment_chain:1"
misc_feature 163328. .168724
/note="assembly_fragment:00049"
misc_feature 168825. .202208
/note="assembly_fragment:00103
clone_end:SP6
vector_side:right"
ORIGIN

Query Match 5.0%; Score 59.6; DB 14; Length 202208;
Best Local Similarity 44.5%; Pred. No. 0.26;
Matches 277; Conservative 0; Mismatches 344; Indels 1; Gaps 1;

```
Qy         113 TTTTTACTATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTCT 172
               || ||| ||| ||  ||  || ||      || || |||   || ||   |||    |||| |
Db       18069 TTATTATTATCATTATATTGTATTTATTATTATTATTATTATTATTATTATTATTGT 18010

Qy         173 CACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTT 232
                 | |   ||  |||      ||||||          | |  |||| |  |         ||
Db       18009 TA-CTACTACTATTATTAAAATTATTATTATTATTACTATTATTATTATTATTATTA 17951

Qy         233 TCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAG 292
               | ||     | || ||| | |   ||| ||     | ||   | | ||||   |  ||
Db       17950 TTATTATTATTGTTATTATTACCATATTATTATTACTATTTATTATTATTATTATTATTA 17891

Qy         293 TTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAA 352
               ||   |  || ||| || | ||  |||   ||   |  ||   |||| ||| | |  |
Db       17890 TTATTTTTATTATTATTATTATTACCATTATTATTATTATTATTGTTGCTATTATTG 17831
```

```
Qy        353 TGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAA 412
              |   |   |   | |||     |  |       || | |   |  | |           | |
Db      17830 TTGTTATTGTTATTATTGTTATTGTTATTATTGTTAATACTACTACTATTATTATTATTA 17771

Qy        413 ACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCA 472
              | |||| || |     |   ||  |    | ||      | || |  |  | | | |
Db      17770 TTATTATTATTACCATCATTATTATTATTATTGTTATTATTATTACTATTATTATTATTA 17711

Qy        473 ATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGAT 532
              |     |||  | |   |   | ||||  |   |||      | ||
Db      17710 TTACCATTATTATTATTATTATTATTATTTTTATTATTATTTTTATTATTATTACTATTA 17651

Qy        533 ATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTC 592
              ||  |   |    ||  |||||||||| | || |  | ||||    | |   || | |
Db      17650 TTATTATTATTACCATTATTATTATTATTATTATTATTATTATTACCATTATTATTA 17591

Qy        593 TTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATC 652
              ||| ||         ||  || ||  |    |||||       ||  ||| ||| | | |
Db      17590 TTATTATTAATATTATTATTATTATTATTATTATTATTATTATTATTATTACCATTA 17531

Qy        653 GGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTT 712
              |  ||  || ||  |   || | | |  || | || |||||     || |    |||
Db      17530 TTATTATTATTATTATTATCATTATTATTATTATTATTATTATTATTATTATTATTATTA 17471

Qy        713 CTACTAATATTTTTGACAACTA 734
              || || |||| |  |||| ||
Db      17470 TTATTATTATTATATACAATTA 17449
```

RESULT 14
AC098822

```
LOCUS       AC098822               181792 bp    DNA     linear   PRI 22-FEB-2002
DEFINITION  Homo sapiens BAC clone RP11-338B23 from 4, complete sequence.
ACCESSION   AC098822 AC025095
VERSION     AC098822.3  GI:18464302
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 181792)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 181792)
  AUTHORS   Ali,J., Haakenson,W. and Creason,K.
  TITLE     The sequence of Homo sapiens BAC clone RP11-338B23
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 181792)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 181792)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 181792)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 1, 2002 this sequence version replaced gi:17977471.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0338B23
                Drafting Center: WIBR
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-631I21, 2000 bp overlap;
            the clone sequenced to the right is RP11-660B13, 2000 bp overlap.
            Actual start of this clone is at base position 161305 of
            RP11-631I21; actual end is at base position 181792 of RP11-338B23.

            There is a region covered by PCR only from 20829 to 20858.  There
            is a simple-sequence repeat from 83554 to 85256.

            The sequence of AC025095 has been incorporated into AC098822.
FEATURES             Location/Qualifiers
     source          1. .181792
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-338B23"
                     /clone_lib="RPCI-11"
     repeat_region   737. .816
                     /rpt_family="MIR"
     repeat_region   1478. .1769
                     /rpt_family="L2"
     repeat_region   2124. .2250
                     /rpt_family="MIR"
     repeat_region   3056. .3349
                     /rpt_family="L1"
     repeat_region   3185. .3206
                     /rpt_family="AT_rich"
     repeat_region   3350. .3763
                     /rpt_family="L1"
     repeat_region   3743. .3782
                     /rpt_family="AT_rich"
     repeat_region   3764. .3790
                     /rpt_family="L1"
     repeat_region   3853. .3956
                     /rpt_family="L1"
     repeat_region   3971. .4044
                     /rpt_family="L1"
```

```
repeat_region   4009. .4031
                /rpt_family="AT_rich"
repeat_region   4056. .4130
                /rpt_family="ERVL"
repeat_region   4276. .4349
                /rpt_family="ERVL"
repeat_region   4353. .4908
                /rpt_family="L1"
repeat_region   5076. .5736
                /rpt_family="L1"
repeat_region   5737. .6109
                /rpt_family="MaLR"
repeat_region   6110. .6569
                /rpt_family="L1"
repeat_region   6594. .6907
                /rpt_family="Alu"
repeat_region   6876. .6907
                /rpt_family="(A)n"
repeat_region   6910. .7076
                /rpt_family="L1"
repeat_region   7077. .7376
                /rpt_family="Alu"
repeat_region   7377. .8198
                /rpt_family="L1"
repeat_region   8201. .8238
                /rpt_family="AT_rich"
misc_feature    8522. .8729
                /note="match to EST BI005753 (NID:g14409827)"
repeat_region   8701. .8813
                /rpt_family="MER1_type?"
repeat_region   8899. .9178
                /rpt_family="MER1_type?"
repeat_region   9264. .9415
                /rpt_family="CT-rich"
repeat_region   9408. .9713
                /rpt_family="Alu"
repeat_region   9570. .9594
                /rpt_family="(T)n"
repeat_region   9718. .10701
                /rpt_family="L1"
misc_feature    10384. .10877
                /note="match to EST AI243631 (NID:g3839028) qh88a07.x1"
repeat_region   10918. .10952
                /rpt_family="AT_rich"
repeat_region   11758. .11794
                /rpt_family="AT_rich"
repeat_region   11940. .12151
                /rpt_family="MIR"
repeat_region   12171. .12195
                /rpt_family="(TTA)n"
repeat_region   12391. .12419
                /rpt_family="AT_rich"
repeat_region   13031. .13175
                /rpt_family="L2"
repeat_region   13761. .14012
                /rpt_family="MaLR"
repeat_region   14013. .14050
                /rpt_family="ERVL"
repeat_region   14051. .14075
                /rpt_family="(T)n"
repeat_region   14361. .14556
                /rpt_family="ERVL"
repeat_region   14546. .14662
                /rpt_family="ERVL"
misc_feature    14645. .14989
                /note="match to EST AA759171 (NID:g2807034) ah70g03.s1"
repeat_region   14679. .14707
                /rpt_family="(TC)n"
repeat_region   14707. .14760
                /rpt_family="(TG)n"
repeat_region   15020. .15241
                /rpt_family="MaLR"
repeat_region   15633. .15658
                /rpt_family="(TTTTA)n"
repeat_region   16504. .16557
                /rpt_family="(CA)n"
repeat_region   16826. .16853
                /rpt_family="AT_rich"
repeat_region   17778. .20489
                /rpt_family="L1"
repeat_region   20470. .20853
                /rpt_family="L1"
repeat_region   20854. .21780
                /rpt_family="L1"
repeat_region   21761. .21817
```

Query Match 4.9%; Score 59; DB 8; Length 181792;
Best Local Similarity 43.5%; Pred. No. 0.34;
Matches 363; Conservative 0; Mismatches 470; Indels 2; Gaps 2;

```
Qy      274 TTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATA 333
Db    83951 TAATATATTTTATAAATGTATATTTATATAATATATAATATATTTAATACATATATATAA 84010

Qy      334 TCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCAT 393
Db    84011 TATATTTAATAAATATATATATTTATAAAATATATTATATAAAATATATATTTTATATAT 84070

Qy      394 TATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGC 453
Db    84071 TATATATTTATAAATATATAATATATAAAATATATATTTTATATATTATATTTTATAAAT 84130

Qy      454 GGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATG 513
Db    84131 ATATATTTATAAATATATAATATATAAAATATATATTTTATATATTATATTTTATAAATA 84190

Qy      514 GAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTT 573
Db    84191 TATATATTTTTATAAATATATAATATATAAATATATTATATGATATATATATTATATATT 84250

Qy      574 ATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCA 633
Db    84251 ATATTATATAATATATAAATATATATTATATATATTTATATATTATATTATATAATAT 84310

Qy      634 CTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATT 693
Db    84311 ATAAATATATATTATATATATTATATTATATAATATATAAATTATATATTATAATATAAT 84370

Qy      694 ATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGATCGGC 753
Db    84371 ATATAATATATATATTTATATATTATATTATATAAATATTATATAAATATAATATTTATA 84430

Qy      754 AGGTTTACAGATTCGTCTCAGGCATCGCGTGGAT-CTACACAGGGGCGGATAGATATGGC 812
Db    84431 TAATATAATATATGTAATATATATTATGATATATAATAAGTAATATATATTATATATAAT 84490

Qy      813 AATTAATTCATTAAACTTCCTGTCAGAACATCCATCAGGTATAGGTCTGGGTACTCAAGG 872
Db    84491 ATATAATAATATATAATAAATGCTATAATATATATTATATAAATAAAATATAAATATATA 84550

Qy      873 TTCAGGAAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGATAATTATTTTTTCTG 932
Db    84551 TAATATATA-ATATATTATATATAAGATATTAATATAAATATATTATATAAATATATTTA 84609

Qy      933 GATCGCCCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCTGGCAAGTCA 992
Db    84610 TATAATATATAATACATATAATATATAATAATATGTAATACTATATAATATAATATATAA 84669

Qy      993 ATTTTATTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAGCAATATGCACTA 1052
Db    84670 TATATATTATAATATATAATAAATAATATAATATATAATATATATTATATAAATAAAATA 84729

Qy     1053 TAGATTATATTTTCTCTTTGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCA 1107
Db    84730 TATATAATATATAATATATTAATATATATAATATATAAGATATATGATACATTAA 84784
```

```
RESULT 15
AX345156
LOCUS       AX345156               18154 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 227 from Patent WO0200928.
ACCESSION   AX345156
VERSION     AX345156.1  GI:18493042
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 227 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .18154
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             4.9%;  Score 58.6;  DB 6;  Length 18154;
  Best Local Similarity   48.4%;  Pred. No. 0.91;
  Matches  193;  Conservative    0;  Mismatches  204;  Indels    2;  Gaps    1;

Qy        350 TAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAAT 409
              |||   | | |  | | || |  ||  | | ||||  |  | |||    |  |  |  |
Db      13388 TAAAATAGGAATGGATAATGATTTTTATTGAATATTTTTTGTATTTATGAAAATGATTAA 13447

Qy        410 TAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTC 469
              | |  ||   |  ||  ||       |    |  ||| |  |    ||||||  |  |  |
Db      13448 TTATTTTTTATTTTAAATAATTATTGTATTTTTGGGATTAATATTGTTTTATTTTTTTTA 13507

Qy        470 TCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAG 529
              | |||| | ||  |||| | |   ||  |  ||| |||| |    | |  ||||  |||
Db      13508 TAAATTGTTGAATTATGTTTAATATATTTTATTTATTATTTTAAAATTTGTTTTTAGAAG 13567

Qy        530 GATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGA 589
              |  || ||| |||||  |  || || ||  |||  || ||    ||  | |    |
Db      13568 TAATTATAAGTTTATTTTTATTTTTTTTTGTTTTGATGTTAAAATTATATTTGTTTTATA 13627

Qy        590 GTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAA 649
              | | | | | ||   |  || ||  |    ||||| |     |||||||||  | |
Db      13628 AAATGATTTGGGTAGTTTTTTTTTATTTTTTTTTTTTATTAAAATTATTTATATAAGATTAG 13687

Qy        650 ATCGGAAGATGCTTT--TTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGC 707
              |       ||  |||    | | |   || ||| |||||||||  |  |      || | |
Db      13688 AATTATTTATTGTTTGAAGTTTTGGTAAAATTTGTTTGTAATATTATTTAGTTTTGTTGT 13747

Qy        708 AATTTCTACTAATATTTTTGACAACTATACAGAAATTTT 746
                |||    |   |  |||   ||   |  |  ||| ||
Db      13748 TTTTTGAGGTGGGAGATTTTTGAATATTGAATTAATATT 13786


Search completed: May 11, 2006, 19:14:56
Job time : 4325.33 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 07:09:59 ; Search time 546.783 Seconds
(without alignments)
14626.705 Million cell updates/sec

Title: US-09-423-093B-3_COPY_5114_6313
Perfect score: 1200
Sequence: 1 atgctgccatttccaccagg.....atagacggtgcacgcgataa 1200

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1200	100.0	12441	2	AAX06750	Aax06750 Salmonell
c	2	64.6	5.4	2000	11	ACL35887	Acl35887 Rice stre
	3	58.6	4.9	18154	6	ABL32254	Abl32254 Human imm
	4	53.4	4.5	16579	10	ADB54118	Adb54118 Pretreate
	5	53.4	4.5	16579	10	ADB54246	Adb54246 Pretreate
	6	53.4	4.5	16579	10	ADE37773	Ade37773 Human che
	7	53.4	4.5	16579	10	ADE37763	Ade37763 Human che
	8	53.4	4.5	16579	13	ADS89274	Ads89274 Oligonucl
	9	53.4	4.5	16579	13	ADS89548	Ads89548 Oligonucl
	10	53.4	4.5	16750	4	AAS46314	Aas46314 Tumour su
	11	53.4	4.5	16750	6	ABL32521	Abl32521 Human imm
	12	51.4	4.3	16373	6	ABL32618	Abl32618 Human imm
	13	51.4	4.3	16373	6	AAD28382	Aad28382 Human che
c	14	50.6	4.2	2000	11	ACL37108	Acl37108 Rice stre
	15	50.6	4.2	6161	6	ABK31299	Abk31299 Signal tr
	16	50.6	4.2	6161	6	ABL70278	Abl70278 Chemicall
	17	50.6	4.2	6161	6	AAS61201	Aas61201 Human gen
	18	50.6	4.2	16766	6	ABL34157	Abl34157 Human imm
	19	50	4.2	5536	6	ABL32179	Abl32179 Human imm
	20	49.6	4.1	6381	6	ABL32706	Abl32706 Human imm
	21	49.2	4.1	2000	11	ACL35363	Acl35363 Rice stre
	22	49.2	4.1	4985	6	ABQ75107	Abq75107 Anopheles
	23	49.2	4.1	4985	10	ACF79720	Acf79720 Mosquito
	24	49.2	4.1	6220	6	ABL33300	Abl33300 Human imm
	25	49.2	4.1	7319	6	ABL34045	Abl34045 Human imm
	26	49.2	4.1	8222	8	ACF62816	Acf62816 Colon can
	27	49.2	4.1	8222	8	ACF62794	Acf62794 Colon can
	28	49.2	4.1	8771	6	ABL33824	Abl33824 Human imm
	29	49.2	4.1	9504	6	ABK28407	Abk28407 DNA trans
	30	49.2	4.1	11222	10	ADB54190	Adb54190 Pretreate
	31	49.2	4.1	11222	10	ADB54318	Adb54318 Pretreate
	32	48.8	4.1	1162	14	ADZ71022	Adz71022 Human chr
	33	48.8	4.1	7317	4	AAS45342	Aas45342 Chemicall
	34	48.8	4.1	7317	6	ABK28173	Abk28173 DNA trans
	35	48.8	4.1	8260	6	ABK28341	Abk28341 DNA trans
	36	48.8	4.1	13202	6	ABL33485	Abl33485 Human imm
c	37	48.6	4.0	1169	14	ADZ71002	Adz71002 Human chr
	38	48.6	4.0	6057	6	ABK31397	Abk31397 Signal tr
	39	48.6	4.0	6057	6	ABL70362	Abl70362 Chemicall
	40	48.6	4.0	8845	4	AAS46544	Aas46544 Tumour su
	41	48.6	4.0	11394	6	ABK28221	Abk28221 DNA trans
c	42	48.6	4.0	110000	12	ADQ97050_0	Adq97050 Human can
	43	48.4	4.0	3051	8	ACF62818	Acf62818 Colon can
	44	48.4	4.0	5952	10	ADB54320	Adb54320 Pretreate
	45	48.4	4.0	8962	6	ABL32686	Abl32686 Human imm

ALIGNMENTS

RESULT 1
AAX06750
ID AAX06750 standard; DNA; 12441 BP.
XX
AC AAX06750;
XX
DT 26-APR-1999 (first entry)
XX
DE Salmonella enterica serovar C2 O antigen gene cluster.
XX
KW O antigen; diagnosis; wbaR gene; wzx gene; wbaL gene; wbaQ gene;
KW wxy gene; wbaW gene; wbaZ gene; transferase; polymerase; flippase;
KW diarrhoea; haemorrhagic colitis; ds.
XX
OS Salmonella enterica.
XX
FH Key Location/Qualifiers
FT CDS 138. .1019
FT /*tag= a
FT /note= "abe gene; encodes AAW88324"
FT CDS 1019. .2359
FT /*tag= b
FT /product= "flippase"
FT /note= "wzx gene, this region is specifically claimed in
FT Claim 9; encodes AAW88325"
FT CDS 2362. .3314
FT /*tag= c
FT /product= "abequosyl_transferase"
FT /note= "wbaR gene, this region is specifically claimed in
FT Claim 9; encodes AAW88326"
FT CDS 3361. .3975
FT /*tag= d
FT /product= "acetyl_transferase"
FT /note= "wbaL gene, this region is specifically claimed in
FT Claim 9; encodes AAW88327"
FT CDS 3977. .5020
FT /*tag= e
FT /product= "rhamnosyl_transferase"
FT /note= "wbaQ gene, this region is specifically claimed in
FT Claim 9; encodes AAW88328"
FT CDS 5114. .6313
FT /*tag= f

```
FT                   /product= "O_antigen_polymerase"
FT                   /note= "wzy gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88329"
FT   CDS             6313. .7323
FT                   /*tag= g
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88330"
FT   CDS             7310. .8467
FT                   /*tag= h
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaW gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88331"
FT   CDS             8482. .9903
FT                   /*tag= i
FT                   /note= "manC gene; encodes AAW88332"
FT   CDS             9911. .11347
FT                   /*tag= j
FT                   /note= "manB gene; encodes AAW88333"
FT   CDS             11426. .12441
FT                   /*tag= k
FT                   /note= "wbaP gene; encodes AAW88334"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88324, AAW88325, AAW88326, AAW88327, AAW88328, AAW88329,
DR   AAW88330, AAW88331, AAW88332, AAW88333, AAW88334.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 9; Fig 9; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Salmonella enterica serotype C2 (strain M67) O
CC   antigen. 11 Open reading frames within the gene cluster were identified
CC   on the basis of homology to known sequences. The genes encode proteins
CC   (see AAW88323-34) involved the synthesis of sugars present in the
CC   polysaccharide antigen, and in the transport or processing of
CC   polysaccharide or oligosaccharide units. The use of nucleic acid
CC   molecules derived from assembly and transport genes, particularly wbd
CC   (transferase), wzx (flippase) and wzy (polymerase) genes, within O
CC   antigen gene clusters improves the specificity of methods for the
CC   detection and identification of O antigens, e.g. in testing food- or
CC   faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of S. enterica
XX
SQ   Sequence 12441 BP; 3887 A; 1876 C; 2518 G; 4160 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1200;  DB 2;  Length 12441;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-257;
  Matches 1200;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCTGCCATTTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5114 ATGCTGCCATTTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCG 5173

Qy         61 TTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5174 TTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACT 5233

Qy        121 ATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5234 ATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGAT 5293

Qy        181 AAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5294 AAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCA 5353

Qy        241 TTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5354 TTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTAC 5413

Qy        301 ATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5414 ATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGA 5473

Qy        361 GAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5474 GAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATT 5533

Qy        421 AATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5534 AATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGA 5593

Qy        481 TATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5594 TATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGA 5653

Qy        541 TTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5654 TTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGA 5713

Qy        601 GGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5714 GGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATG 5773

Qy        661 CTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5774 CTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAAT 5833

Qy        721 ATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5834 ATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCG 5893

Qy        781 CGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAGAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5894 CGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAGAA 5953

Qy        841 CATCCATCAGGTATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5954 CATCCATCAGGTATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGAT 6013

Qy        901 AATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6014 AATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGC 6073

Qy        961 TTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6074 TTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGA 6133

Qy       1021 ATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6134 ATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATA 6193

Qy       1081 TATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6194 TATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTAT 6253

Qy       1141 TGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGACGGTGCACGCGATAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6254 TGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGACGGTGCACGCGATAA 6313
```

```
RESULT 2
ACL35887/c
ID   ACL35887 standard; cDNA; 2000 BP.
XX
AC   ACL35887;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Rice stress-regulated promoter SEQ ID NO:14450.
XX
KW   ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW   agriculture.
XX
OS   Oryza sativa.
XX
PN   WO2003008540-A2.
XX
PD   30-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US019668.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   24-AUG-2001; 2001US-0314662P.
PR   26-SEP-2001; 2001US-0325277P.
PR   21-NOV-2001; 2001US-0332132P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Kreps J,  Briggs SP,  Cooper B,  Glazebrook J,  Goff SA,  Katagiri F;
PI   Moughamer T,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-248011/24.
XX
PT   New stress-responsive nucleic acid, useful for altering the
PT   responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT   stress, salt stress or osmotic stress.
XX
PS   Claim 48; SEQ ID NO 14450; 89pp; English.
XX
CC   The invention relates to novel abiotic stress responsive polynucleotides
CC   and polypeptides. Also disclosed are vectors, expression cassettes, host
CC   cells, and plants containing such polynucleotides. Also disclosed are
CC   methods for using the polynucleotides and polypeptides to alter the
CC   responsiveness of a plant to abiotic stress. The invention is useful in
CC   agriculture. The nucleic acid is useful for determining whether a test
CC   plant has been exposed to an abiotic stress condition. It is also useful
CC   for selecting an agent that alters abiotic stress regulated
CC   polynucleotide expression in a plant cell, and to identify a homolog or
CC   ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC   molecule and the polypeptide encoded by it are useful in altering the
CC   responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC   stress, osmotic stress or any of their combinations. The present sequence
CC   is used in the exemplification of the invention
XX
SQ   Sequence 2000 BP; 410 A; 329 C; 253 G; 416 T; 0 U; 592 Other;

  Query Match             5.4%;  Score 64.6;  DB 11;  Length 2000;
  Best Local Similarity   16.6%;  Pred. No. 0.00017;
  Matches  155;  Conservative  345;  Mismatches  415;  Indels   18;  Gaps    3;

Qy        140 TAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGATAAAATTCAAGCAATTATGG 199
              ||||| |:|:  |:||: :  :::: :::: | :    :: ::: ||::   :::| ::
Db        950 TAATAATWTRMYTMATMTMAMWWMYSRWMYRATRMTAWAMWTMWWCTCRWARYMKWASSK 891

Qy        200 GGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTAT 259
              : |   ::: |: : :  |:::    :  |:::| :     ::      |::|:| :||
Db        890 RTGYTWRRWWTWYWCYSRTMKWAWWAKGTGYKWTTRRYTWKYKSYGATSTWWAWATRTAR 831

Qy        260 CA------TACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAA 313
              |:      | :  |       | : |||: :|::|:|  | :   :  ::|:::::| :::
Db        830 CWAWKWRCTGMGTGKTKATKTAWTATAWRKAWMAMARTGCWRMKKGWWYAWWWWWTWYWR 771

Qy        314 TGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAAT 373
              : |:|:  :: ::: :  : :: :| : ||  :|:| :|   | ::: :    : :   ::
Db        770 GKATWAYMMWKYMRMTSRMMYRWARKTTTCYKGWATKATMMAKWSRSKSMSMYGKMSGWY 711

Qy        374 TGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATTAATAGC------- 426
              :::|   : ||:::|::  :   :| | ::: :  ::  |: : :|:  :
Db        710 YSSTKWMWGATRMWGMKWGKGCRRCGGYMKWSMKSKWTCASAWGMTKYKYSYRCRGCGGT 651

Qy        427 --TATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGATATA 484
                |:: :       ::   |::::|:| :: :||:    :: :  :   |  |: :: |
Db        650 WTTRWWMRAWMCMAKWTKTGSYYWCRGYSKYKTTWYWRAWWWSCWKKTGTGRTKYMWGTS 591

Qy        485 TGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGATTAT 544
              |  |  :     :  : :::: :|::|||   :|:| ::| ::   : ::::    |
Db        590 TYGTTKMYSAYWKRMWSKKRWGRTKYATGWKYWGYTGWYCAYWSAAMCWKMRKGTKTRTG 531

Qy        545 TGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGAGGAG 604
              :  ||:  :  |::: : :    ::| |   : |:  |:: |: :: :: | | :
Db        530 RKSRGCWRTKCKTYWRMYRWMGYRWRTGTSRGRGGMAATKMSTKYRWKRWWRMAWATKSW 471

Qy        605 CAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTT 664
              :: :::|::   : :: ::: :|:|   |:   :   || |   :  |::  | :
Db        470 KWRGMMWTKWCTKYAYKRWYMRAKAYGKKTTWGCSTYGKCTTWAYTRASYGAKWCTRTGK 411

Qy        665 TTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAATATTT 724
              ||:|| : :: ::|| | |: : :|  : |   : :  : : ::::|   :|:|:::|:
Db        410 TTKGTCRGMKYRMTTAAGTKAKYYKAKRKTTTKTMWKMRYGSKWRWKTTAGYTRAKWKTY 351

Qy        725 TTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCGCGTG 784
              :: :    ::    : :: ::: :| | :    ::: : ::: :|    :   | :: |
Db        350 KYTRTTSWWRSKWKKTWWRKKWWRTKGTYKCTGKWRTMAMWWKKTARRTSWWKTKKYTTM 291

Qy        785 GATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAGAACATC 844
              : : : ||:: | : :    ::: :  |  |:   : :    ::: |||    |  |:
Db        290 KGKRWKCAMRAGTKAKWTCTRWWYYMYWAKRTRTGMMCKRSKGMWWWCTGAAYCACWAYA 231

Qy        845 CATCAGGTATAGGTC---TGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGATA 901
              || :  | : :  :    | | :  ::|:  :: ::   |:: :| :|:  : : : : :
Db        230 KATGWYCTKYGRCAMACWTTGTYKTWSARWWYWTVKWTGAYWSKCAWTSWYWYWKMWWRM 171

Qy        902 ATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGCT 961
              |:||:| | : ::   ::   :: :|:::: | |  :  :  |   :: ::   :|  ::
Db        170 AWAGRTGATWRWMAATYRCACRKCWTKWMWMGGTGAYAAWAAATTRKRMKRMATWTMTYW 111

Qy        962 TAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAA 1021
              : :: ::::|| : :: : :    :: |: :: |::::: :  ||   :|  ::    |
Db        110 KKMYRMWWWATTAYKWYAMAYAYATWWWTMTRYATWWWWMMYTCACACGWCATRWRYATA 51

Qy       1022 TATATGGCAGTCATTGTAGCAATATGCACTATA 1054
              |::||  | : :  :    :|::|  :: :: |
Db         50 TWWATWACGRCYYAYAMGWSAWKARCMWAKWAA 18


RESULT 3
ABL32254
ID   ABL32254 standard; DNA; 18154 BP.
XX
AC   ABL32254;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 227.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
```

XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 227; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 18154 BP; 5274 A; 112 C; 3279 G; 9489 T; 0 U; 0 Other;

Query Match 4.9%; Score 58.6; DB 6; Length 18154;
Best Local Similarity 48.4%; Pred. No. 0.0058;
Matches 193; Conservative 0; Mismatches 204; Indels 2; Gaps 1;

Qy 350 TAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAAT 409
Db 13388 TAAAATAGGAATGGATAATGATTTTTATTGAATATTTTTTGTATTTATGAAAATGATTAA 13447

Qy 410 TAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTC 469
Db 13448 TTATTTTTTATTTTAAATAATTATTGTATTTTTGGGATTAATATTGTTTTATTTTTTTTA 13507

Qy 470 TCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAG 529
Db 13508 TAAATTGTTGAATTATGTTTAATATATTTTATTTATTATTTTAAAATTTGTTTTTAGAAG 13567

Qy 530 GATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGA 589
Db 13568 TAATTATAAGTTTATTTTTATTTTTTTTTGTTTTGATGTTAAAATTATATTTGTTTTATA 13627

Qy 590 GTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAA 649
Db 13628 AAATGATTTGGGTAGTTTTTTTTATTTTTTTTTTTTATTAAAATTATTTATATAAGATTAG 13687

Qy 650 ATCGGAAGATGCTTT--TTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGC 707
Db 13688 AATTATTTATTGTTTGAAGTTTTGGTAAAATTTGTTTGTAATATTATTTAGTTTTGTTGT 13747

Qy 708 AATTTCTACTAATATTTTTGACAACTATACAGAAATTTT 746
Db 13748 TTTTTGAGGTGGGAGATTTTTGAATATTGAATTAATATT 13786

RESULT 4
ADB54118
ID ADB54118 standard; DNA; 16579 BP.
XX
AC ADB54118;
XX
DT 04-DEC-2003 (first entry)

XX
DE Pretreated genomic DNA region 42.
XX
KW colon cell proliferative disorder; non methylated CpG dinucleotide;
KW cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.
XX
OS Unidentified.
XX
PN WO2003072821-A2.
XX
PD 04-SEP-2003.
XX
PF 27-FEB-2003; 2003WO-EP002035.
XX
PR 27-FEB-2002; 2002EP-00004551.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Adorjan P, Burger M, Maier S, Nimmrich I, Becker E, Lesche R;
PI Rujan T, Schmitt A;
XX
DR WPI; 2003-731620/69.
XX
PT Detecting and differentiating between colon cell proliferative disorders
PT associated with a gene or its regulatory regions comprises contacting a
PT target nucleic acid in a biological sample obtained from the subject with
PT a reagent.
XX
PS Claim 32; SEQ ID NO 174; 74pp; English.
XX
CC The invention relates to a novel method for detecting and differentiating
CC between colon cell proliferative disorders associated with at least one
CC gene or its regulatory regions. The method comprises contacting a target
CC nucleic acid in a biological sample obtained from the subject with at
CC least one reagent or a series of reagents, where the reagent or series of
CC reagents, distinguishes between methylated and non methylated CpG
CC dinucleotides within the target nucleic acid. The molecules of the
CC invention demonstrate cytostatic activity whilst the method may useful
CC for detecting and differentiating between colon cell proliferative
CC disorders, including cancers such as colon adenoma and colon carcinoma.
CC The PNA (peptide nucleic acid)-oligomers are useful as probes for
CC determining cytosine methylation state or single nucleotide
CC polymorphisms. The current sequence is that of the pretreated genomic DNA
CC region of the invention. This sequence is not shown within the
CC specification but is taken from Wipoweb.
XX
SQ Sequence 16579 BP; 4923 A; 274 C; 3197 G; 8185 T; 0 U; 0 Other;

Query Match 4.5%; Score 53.4; DB 10; Length 16579;
Best Local Similarity 42.1%; Pred. No. 0.083;
Matches 303; Conservative 0; Mismatches 416; Indels 0; Gaps 0;

Qy 30 AAGGGATGTACTCAATGTATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAG 89
Db 13377 AAGTTATATAATTAATAGTTTAGTGGTTTTTTTAGTGTTAATTATAGTAATTTTAGATTT 13436

Qy 90 GAAAAAAACTTATTTCCCGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATG 149
Db 13437 AAGATTGTTTGATTTAGGAATGGAGAGAAATAAAATAAAATGAATGTTTATTGAAATATA 13496

Qy 150 GGTAATAGCGTTAACTATATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAG 209
Db 13497 TATTTTAGATATTAAGATAGATTTTTTTGTGAAATATTTTAAGATGTGGATTTGTATTTT 13556

Qy 210 TTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCA 269
Db 13557 TATTTTTAAATTTGAGTAAATTTAGTTGAGTTTTAAAGAGTTTAAGTAATTTGTTTAAGG 13616

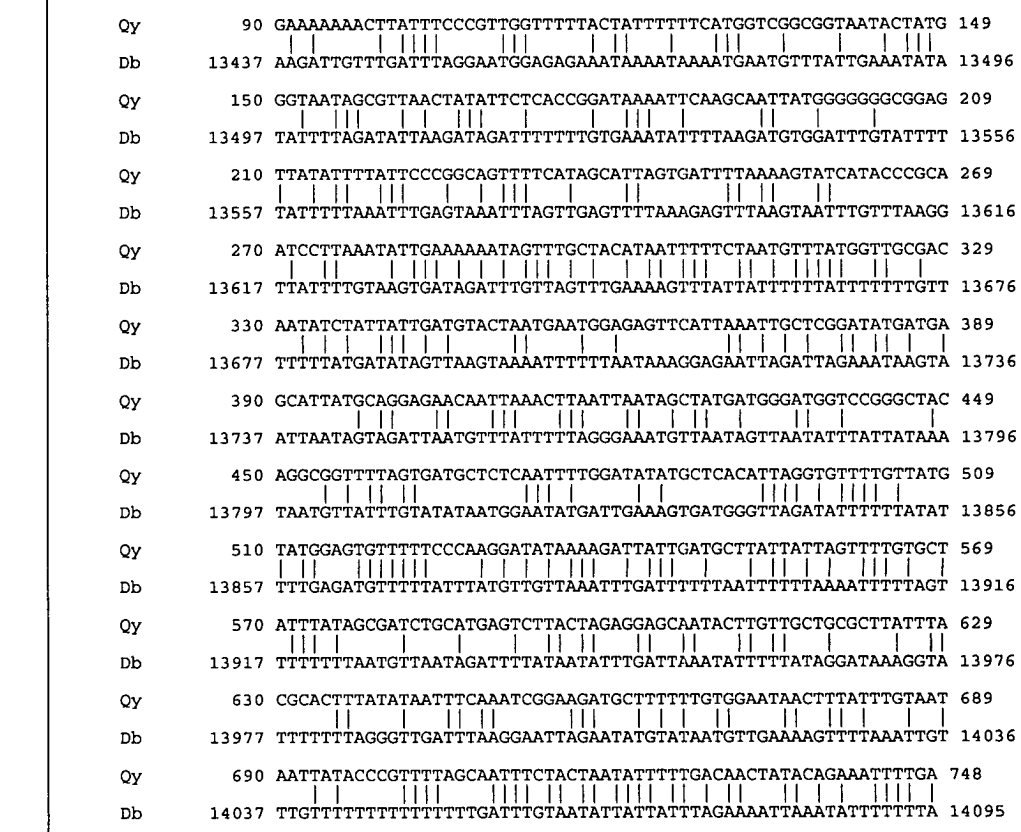
Qy 270 ATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGAC 329
Db 13617 TTATTTTGTAAGTGATAGATTTGTTAGTTTGAAAAGTTTATTATTTTTTATTTTTTTGTT 13676

Qy 330 AATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGA 389



RESULT 5
ADB54246
ID ADB54246 standard; DNA; 16579 BP.
XX
AC ADB54246;
XX
DT 04-DEC-2003 (first entry)
XX
DE Pretreated genomic DNA region 170.
XX
KW colon cell proliferative disorder; non methylated CpG dinucleotide;
KW cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.
XX
OS Unidentified.
XX
PN WO2003072821-A2.
XX
PD 04-SEP-2003.
XX
PF 27-FEB-2003; 2003WO-EP002035.
XX
PR 27-FEB-2002; 2002EP-00004551.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Adorjan P, Burger M, Maier S, Nimmrich I, Becker E, Lesche R;
PI Rujan T, Schmitt A;
XX
DR WPI; 2003-731620/69.
XX
PT Detecting and differentiating between colon cell proliferative disorders
PT associated with a gene or its regulatory regions comprises contacting a
PT target nucleic acid in a biological sample obtained from the subject with
PT a reagent.
XX
PS Claim 32; SEQ ID NO 302; 74pp; English.
XX
CC The invention relates to a novel method for detecting and differentiating
CC between colon cell proliferative disorders associated with at least one
CC gene or its regulatory regions. The method comprises contacting a target
CC nucleic acid in a biological sample obtained from the subject with at
CC least one reagent or a series of reagents, where the reagent or series of
CC reagents, distinguishes between methylated and non methylated CpG
CC dinucleotides within the target nucleic acid. The molecules of the
CC invention demonstrate cytostatic activity whilst the method may useful
CC for detecting and differentiating between colon cell proliferative
CC disorders, including cancers such as colon adenoma and colon carcinoma.
CC The PNA (peptide nucleic acid)-oligomers are useful as probes for
CC determining cytosine methylation state or single nucleotide
CC polymorphisms. The current sequence is that of the pretreated genomic DNA
CC region of the invention. This sequence is not shown within the
CC specification but is taken from Wipoweb.
XX
SQ Sequence 16579 BP; 4923 A; 0 C; 3197 G; 8459 T; 0 U; 0 Other;

Query Match 4.5%; Score 53.4; DB 10; Length 16579;
Best Local Similarity 42.1%; Pred. No. 0.083;
Matches 303; Conservative 0; Mismatches 416; Indels 0; Gaps 0;



RESULT 6
ADE37773
ID ADE37773 standard; DNA; 16579 BP.
XX
AC ADE37773;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human chemically treated TPEF nucleotide sequence SEQ ID NO:19.
XX

```
KW   chemically pretreated genomic DNA; human; versican; TPEF; H-cadherin;
KW   calcitonin; EYA4; cytostatic; gene therapy;
KW   colon cell proliferative disorder; cytosine methylation state;
KW   single nucleotide polymorphism; SNP; disease analysis; CpG dinucleotide;
KW   gene; ds.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO2003072820-A2.
XX
PD   04-SEP-2003.
XX
PF   27-FEB-2003; 2003WO-EP002034.
XX
PR   27-FEB-2002; 2002EP-00004551.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Adorjan P,  Burger M,  Maier S,  Lesche R,  Cottrell S,  Mooney S;
XX
DR   WPI; 2003-731619/69.
XX
PT   New nucleic acid comprising a sequence of at least 18 bases in length of
PT   a segment of the chemically pretreated genomic DNA, useful for treating
PT   colon cell proliferative disorders.
XX
PS   Claim 1; SEQ ID NO 19; 147pp; English.
XX
CC   The present invention describes a nucleic acid (I) comprising a sequence
CC   of at least 18 bases in length of a segment of the chemically pretreated
CC   genomic DNA of any of the 5 sequences of 965, 16579, 3000, 1984, or 7833
CC   bp, which represent human versican, TPEF, H-cadherin, calcitonin and EYA4
CC   respectively (see ADE37755 to ADE37759), or its complement. Also
CC   described: (1) an oligomer, in particular an oligonucleotide or peptide
CC   nucleic acid (PNA)-oligomer comprising in each case of at least one base
CC   sequence having a length of at least 9 nucleotides which is complementary
CC   to, or hybridises under moderately stringent or stringent conditions to a
CC   pretreated genomic DNA; (2) a set of oligomers comprising at least two of
CC   the oligomer of (1); (3) manufacturing an arrangement of different
CC   oligomers (array) fixed to a carrier material; (4) an array of different
CC   oligonucleotide- and/or PNA-oligomer sequences, which are arranged on a
CC   plane solid phase in the form of a rectangular or hexagonal lattice; (5)
CC   a composition of matter comprising the nucleic acid and a buffer
CC   comprising 1-5 mM magnesium chloride, 100-500 micromole dNTP, 0.5-5 units
CC   of taq polymerase, and the oligomer; (6) detecting, differentiating or
CC   distinguishing between colon cell proliferative disorders; and (7)
CC   detecting a colon cell proliferative disorder. (I) has cytostatic
CC   activity, and can be used in gene therapy. The versican, TPEF, H-
CC   cadherin, calcitonin and EYA4 genes, and the polypeptides expressed by
CC   them, can be used for detecting, differentiating or distinguishing
CC   between colon cell proliferative disorders. The oligomers are useful for
CC   detecting the cytosine methylation state and/or single nucleotide
CC   polymorphisms (SNPs) within nucleic acid sequences. The array is useful
CC   for analysing diseases associated with the methylation state of the CpG
CC   dinucleotides. The present sequence is used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 16579 BP; 4923 A; 0 C; 3197 G; 8459 T; 0 U; 0 Other;

  Query Match             4.5%;  Score 53.4;  DB 10;  Length 16579;
  Best Local Similarity   42.1%;  Pred. No. 0.083;
  Matches  303;  Conservative    0;  Mismatches  416;  Indels    0;  Gaps    0;

Qy         30 AAGGGATGTACTCAATGTATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAG 89
              |||  || || | |||    ||  | ||    |||||| || ||  | |  |  | |||
Db      13377 AAGTTATATAATTAATAGTTTAGTGGTTTTTTTAGTGTTAATTATAGTAATTTTAGATTT 13436

Qy         90 GAAAAAAACTTATTTCCCGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATG 149
                | |      | ||||     |||     | ||    |   |||    |    |   | |||
Db      13437 AAGATTGTTTGATTTAGGAATGGAGAGAAATAAAATAAAATGAATGTTTATTGAAATATA 13496

Qy        150 GGTAATAGCGTTAACTATATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAG 209
```

```
              |  |||   | |  |||   |      |  |||   |       ||   |    |
Db      13497 TATTTTAGATATTAAGATAGATTTTTTTGTGAAATATTTTAAGATGTGGATTTGTATTTT 13556

Qy        210 TTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCA 269
              |  | ||  |||    |  | |||   |     ||        || ||   ||
Db      13557 TATTTTTAAATTTGAGTAAATTTAGTTGAGTTTTAAAGAGTTTAAGTAATTTGTTTAAGG 13616

Qy        270 ATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGAC 329
               |  ||     | ||| | | | ||| | |  | || |||  || | |||||  ||  |
Db      13617 TTATTTTGTAAGTGATAGATTTGTTAGTTTGAAAAGTTTATTATTTTTTATTTTTTGTT 13676

Qy        330 AATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGA 389
               | | |  ||| | |     ||    |  |          || | | |  || || |  |
Db      13677 TTTTTATGATATAGTTAAGTAAAATTTTTTAATAAAGGAGAATTAGATTAGAAATAAGTA 13736

Qy        390 GCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTAC 449
                  | ||    ||   |||   |||   ||  | ||  |    ||  |     |
Db      13737 ATTAATAGTAGATTAATGTTTATTTTTAGGGAAATGTTAATAGTTAATATTTATTATAAA 13796

Qy        450 AGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATG 509
               | | || ||          ||| |    | |      |||| | |||| |
Db      13797 TAATGTTATTTGTATATAATGGAATATGATTGAAAGTGATGGGTTAGATATTTTTTATAT 13856

Qy        510 TATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCT 569
               | ||   |||||||     | | | | |||  | |||  |    | || | |  ||| | |
Db      13857 TTTGAGATGTTTTTATTTATGTTGTTAAATTTGATTTTTTAATTTTTTAAAATTTTTAGT 13916

Qy        570 ATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTA 629
               ||| |     |    |    |  || ||   ||  ||   || ||    |     | ||
Db      13917 TTTTTTTAATGTTAATAGATTTTATAATATTTGATTAAATATTTTTATAGGATAAAGGTA 13976

Qy        630 CGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAAT 689
                 ||    |  || ||      |||  | | |  ||   ||  || |    |   |
Db      13977 TTTTTTTAGGGTTGATTTAAGGAATTAGAATATGTATAATGTTGAAAAGTTTTAAATTGT 14036

Qy        690 AATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGA 748
               | |     ||||    |||| || || |||| || | ||   || | |  |||| |
Db      14037 TTGTTTTTTTTTTTTTTTGATTTGTAATATTATTATTTAGAAAATTAAATATTTTTTTA 14095
```

```
RESULT 7
ADE37763
ID   ADE37763 standard; DNA; 16579 BP.
XX
AC   ADE37763;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human chemically treated TPEF nucleotide sequence SEQ ID NO:9.
XX
KW   chemically pretreated genomic DNA; human; versican; TPEF; H-cadherin;
KW   calcitonin; EYA4; cytostatic; gene therapy;
KW   colon cell proliferative disorder; cytosine methylation state;
KW   single nucleotide polymorphism; SNP; disease analysis; CpG dinucleotide;
KW   gene; ds.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO2003072820-A2.
XX
PD   04-SEP-2003.
XX
PF   27-FEB-2003; 2003WO-EP002034.
XX
PR   27-FEB-2002; 2002EP-00004551.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Adorjan P,  Burger M,  Maier S,  Lesche R,  Cottrell S,  Mooney S;
XX
DR   WPI; 2003-731619/69.
XX
```

```
PT   New nucleic acid comprising a sequence of at least 18 bases in length of
PT   a segment of the chemically pretreated genomic DNA, useful for treating
PT   colon cell proliferative disorders.
XX
PS   Claim 1; SEQ ID NO 9; 147pp; English.
XX
CC   The present invention describes a nucleic acid (I) comprising a sequence
CC   of at least 18 bases in length of a segment of the chemically pretreated
CC   genomic DNA of any of the 5 sequences of 965, 16579, 3000, 1984, or 7833
CC   bp, which represent human versican, TPEF, H-cadherin, calcitonin and EYA4
CC   respectively (see ADE37755 to ADE37759), or its complement. Also
CC   described: (1) an oligomer, in particular an oligonucleotide or peptide
CC   nucleic acid (PNA)-oligomer comprising in each case of at least one base
CC   sequence having a length of at least 9 nucleotides which is complementary
CC   to, or hybridises under moderately stringent or stringent conditions to a
CC   pretreated genomic DNA; (2) a set of oligomers comprising at least two of
CC   the oligomer of (1); (3) manufacturing an arrangement of different
CC   oligomers (array) fixed to a carrier material; (4) an array of different
CC   oligonucleotide- and/or PNA-oligomer sequences, which are arranged on a
CC   plane solid phase in the form of a rectangular or hexagonal lattice; (5)
CC   a composition of matter comprising the nucleic acid and a buffer
CC   comprising 1-5 mM magnesium chloride, 100-500 micromole dNTP, 0.5-5 units
CC   of taq polymerase, and the oligomer; (6) detecting, differentiating or
CC   distinguishing between colon cell proliferative disorders; and (7)
CC   detecting a colon cell proliferative disorder. (I) has cytostatic
CC   activity, and can be used in gene therapy. The versican, TPEF, H-
CC   cadherin, calcitonin and EYA4 genes, and the polypeptides expressed by
CC   them, can be used for detecting, differentiating or distinguishing
CC   between colon cell proliferative disorders. The oligomers are useful for
CC   detecting the cytosine methylation state and/or single nucleotide
CC   polymorphisms (SNPs) within nucleic acid sequences. The array is useful
CC   for analysing diseases associated with the methylation state of the CpG
CC   dinucleotides. The present sequence is used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 16579 BP; 4923 A; 274 C; 3197 G; 8185 T; 0 U; 0 Other;

  Query Match             4.5%;  Score 53.4;  DB 10;  Length 16579;
  Best Local Similarity   42.1%;  Pred. No. 0.083;
  Matches  303;  Conservative    0;  Mismatches  416;  Indels    0;  Gaps    0;

Qy         30 AAGGGATGTACTCAATGTATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAG 89
              |||  || || | |||   ||  | ||    ||||| || ||  | |  |  | |||
Db      13377 AAGTTATATAATTAATAGTTTAGTGGTTTTTTTAGTGTTAATTATAGTAATTTTAGATTT 13436

Qy         90 GAAAAAAACTTATTTCCCGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATG 149
               | |      | ||||     |||     | ||   |   |||   |    |   | |||
Db      13437 AAGATTGTTTGATTTAGGAATGGAGAGAAATAAAATAAAATGAATGTTTATTGAAATATA 13496

Qy        150 GGTAATAGCGTTAACTATATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAG 209
                |  |||   | |  |||   |      |  |||  |        ||   |    |
Db      13497 TATTTTAGATATTAAGATAGATTTTTTTGTGAAATATTTTAAGATGTGGATTTGTATTTT 13556

Qy        210 TTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCA 269
              |  | ||  |||   |  | |||   |      ||       || ||   ||
Db      13557 TATTTTTAAATTTGAGTAAATTTAGTTGAGTTTTAAAGAGTTTAAGTAATTTGTTTAAGG 13616

Qy        270 ATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGAC 329
              |  ||    | ||| | | | ||| | |  | || |||  || | |||||  ||  |
Db      13617 TTATTTTGTAAGTGATAGATTTGTTAGTTTGAAAAGTTTATTATTTTTTATTTTTTTGTT 13676

Qy        330 AATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGA 389
              | | |  ||| | |     ||    |  |          || | | |  || || |  |
Db      13677 TTTTTATGATATAGTTAAGTAAAATTTTTTAATAAAGGAGAATTAGATTAGAAATAAGTA 13736

Qy        390 GCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTAC 449
                | ||   ||   |||   |||   ||  | ||  |    ||  |       |
Db      13737 ATTAATAGTAGATTAATGTTTATTTTTAGGGAAATGTTAATAGTTAATATTTATTATAAA 13796

Qy        450 AGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATG 509
                | | || ||         ||| |     | |       |||| | |||| |
Db      13797 TAATGTTATTTGTATATAATGGAATATGATTGAAAGTGATGGGTTAGATATTTTTTATAT 13856

Qy        510 TATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCT 569
              | ||   |||||||    | | | | |||  | |||  |   | || | |  ||| | |
Db      13857 TTTGAGATGTTTTTATTTATGTTGTTAAATTTGATTTTTTAATTTTTTAAAATTTTTAGT 13916

Qy        570 ATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTA 629
               ||| |     |    |    |  || ||   ||  ||   || ||   |     | ||
Db      13917 TTTTTTTAATGTTAATAGATTTTATAATATTTGATTAAATATTTTTATAGGATAAAGGTA 13976

Qy        630 CGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAAT 689
                  ||    |   || ||      |||   | | |  ||    ||  || |   | |
Db      13977 TTTTTTTAGGGTTGATTTAAGGAATTAGAATATGTATAATGTTGAAAAGTTTTAAATTGT 14036

Qy        690 AATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGA 748
                | |      ||||    |||| || || |||| || | ||   || | |  |||| |
Db      14037 TTGTTTTTTTTTTTTTTTGATTTGTAATATTATTATTTAGAAAATTAAATATTTTTTTA 14095


RESULT 8
ADS89274
ID   ADS89274 standard; DNA; 16579 BP.
XX
AC   ADS89274;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Oligonucleotide of the invention SEQ ID NO:290.
XX
KW   ss; cell proliferative disorder; breast; methylation; cytostatic;
KW   gene therapy; single nucleotide polymorphism; SNP.
XX
OS   Unidentified.
XX
PN   WO2004035803-A2.
XX
PD   29-APR-2004.
XX
PF   01-OCT-2003; 2003WO-EP010881.
XX
PR   01-OCT-2002; 2002DE-01045779.
PR   07-JAN-2003; 2003DE-01000096.
PR   17-APR-2003; 2003DE-01017955.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Foekens J,  Harbeck N,  Koenig T,  Maier S,  Martens J,  Model F;
PI   Nimmrich I,  Rujan T,  Schmitt A,  Schmitt M,  Look MP,  Marx A;
XX
DR   WPI; 2004-348468/32.
XX
PT   Predicting responsiveness of a subject with breast cell proliferative
PT   disorder, useful for treating or differentiating breast cell
PT   proliferative disorders comprises analyzing methylation pattern of a
PT   genomic DNA from the subject.
XX
PS   Disclosure; SEQ ID NO 290; 104pp; English.
XX
CC   The invention relates to a novel method for predicting the responsiveness
CC   of a subject with a cell proliferative disorder of the breast tissues to
CC   a therapy comprising analysing the methylation pattern of a target
CC   nucleic acid by contacting at least one of the target nucleic acids in a
CC   biological sample obtained from the subject prior to or during treatment.
CC   The method of the invention has cytostatic activity, and may have a use
CC   in gene therapy. The set of oligonucleotides comprising at least two of
CC   the oligomers are useful for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms (SNPs) within the sequences. The
CC   methods, nucleic acid, oligonucleotide, and kit are useful for the
CC   treatment, characterisation, classification and/or differentiation, of
CC   breast cell proliferative disorders. The method is also useful for
CC   predicting the responsiveness of a subject with a cell proliferative
CC   disorder of the breast tissues to a therapy. The present sequence is used
CC   in the exemplification of the invention.
XX
SQ   Sequence 16579 BP; 4923 A; 274 C; 3197 G; 8185 T; 0 U; 0 Other;
```

```
  Query Match             4.5%;  Score 53.4;  DB 13;  Length 16579;
  Best Local Similarity   42.1%;  Pred. No. 0.083;
  Matches  303;  Conservative    0;  Mismatches  416;  Indels    0;  Gaps    0;

Qy         30 AAGGGATGTACTCAATGTATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAG 89
              |||  || || | |||   ||  | ||    ||||||  || ||  | |  |  | |||
Db      13377 AAGTTATATAATTAATAGTTTAGTGGTTTTTTTAGTGTTAATTATAGTAATTTTAGATTT 13436

Qy         90 GAAAAAAACTTATTTCCCGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATG 149
               | |      | ||||     |||      | ||   |   |||   |    |   | |||
Db      13437 AAGATTGTTTGATTTAGGAATGGAGAGAAATAAAATAAAATGAATGTTTATTGAAATATA 13496

Qy        150 GGTAATAGCGTTAACTATATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAG 209
                |  |||   | |  |||   |      |  |||  |       ||   |   |
Db      13497 TATTTTAGATATTAAGATAGATTTTTTTGTGAAATATTTTAAGATGTGGATTTGTATTTT 13556

Qy        210 TTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCA 269
              | | ||  |||   | | |||   |    ||       || ||   ||
Db      13557 TATTTTTAAATTTGAGTAAATTTAGTTGAGTTTTAAAGAGTTTAAGTAATTTGTTTAAGG 13616

Qy        270 ATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGAC 329
              | ||     | ||| | | | ||| | |  | || |||  || | |||||  ||  |
Db      13617 TTATTTTGTAAGTGATAGATTTGTTAGTTTGAAAAGTTTATTATTTTTTATTTTTTTGTT 13676

Qy        330 AATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGA 389
              | | |  ||| | |    ||   | |       || | | |  || || |  |
Db      13677 TTTTTATGATATAGTTAAGTAAAATTTTTTAATAAAGGAGAATTAGATTAGAAATAAGTA 13736

Qy        390 GCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTAC 449
                | ||   ||   |||   |||   ||  | ||  |    ||  |       |
Db      13737 ATTAATAGTAGATTAATGTTTATTTTTAGGGAAATGTTAATAGTTAATATTTATTATAAA 13796

Qy        450 AGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATG 509
                | | || ||       ||| |     | |         |||| | ||||  |
Db      13797 TAATGTTATTTGTATATAATGGAATATGATTGAAAGTGATGGGTTAGATATTTTTTATAT 13856

Qy        510 TATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCT 569
              | ||   |||||||     | | | | |||  | |||  |    | || | |  ||| |  |
Db      13857 TTTGAGATGTTTTTATTTATGTTGTTAAATTTGATTTTTTAATTTTTTAAAATTTTTAGT 13916

Qy        570 ATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTA 629
               ||| |   |   |   |  || ||  ||  ||  || ||   |     |  ||
Db      13917 TTTTTTTAATGTTAATAGATTTTATAATATTTGATTAAATATTTTTATAGGATAAAGGTA 13976

Qy        630 CGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAAT 689
                 ||   |  || ||       |||  | | |  ||    ||  || |   |  |
Db      13977 TTTTTTTAGGGTTGATTTAAGGAATTAGAATATGTATAATGTTGAAAAGTTTTAAATTGT 14036

Qy        690 AATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGA 748
              | |     ||||    |||| || || |||| || | ||   || | |  |||| |
Db      14037 TTGTTTTTTTTTTTTTTTGATTTGTAATATTATTATTTAGAAAATTAAATATTTTTTTA 14095


RESULT 9
ADS89548
ID   ADS89548 standard; DNA; 16579 BP.
XX
AC   ADS89548;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Oligonucleotide of the invention SEQ ID NO:564.
XX
KW   ss; cell proliferative disorder; breast; methylation; cytostatic;
KW   gene therapy; single nucleotide polymorphism; SNP.
XX
OS   Unidentified.
XX
PN   WO2004035803-A2.
XX
PD   29-APR-2004.
XX
PF   01-OCT-2003; 2003WO-EP010881.
XX
PR   01-OCT-2002; 2002DE-01045779.
PR   07-JAN-2003; 2003DE-01000096.
PR   17-APR-2003; 2003DE-01017955.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Foekens J,  Harbeck N,  Koenig T,  Maier S,  Martens J,  Model F;
PI   Nimmrich I,  Rujan T,  Schmitt A,  Schmitt M,  Look MP,  Marx A;
XX
DR   WPI; 2004-348468/32.
XX
PT   Predicting responsiveness of a subject with breast cell proliferative
PT   disorder, useful for treating or differentiating breast cell
PT   proliferative disorders comprises analyzing methylation pattern of a
PT   genomic DNA from the subject.
XX
PS   Disclosure; SEQ ID NO 564; 104pp; English.
XX
CC   The invention relates to a novel method for predicting the responsiveness
CC   of a subject with a cell proliferative disorder of the breast tissues to
CC   a therapy comprising analysing the methylation pattern of a target
CC   nucleic acid by contacting at least one of the target nucleic acids in a
CC   biological sample obtained from the subject prior to or during treatment.
CC   The method of the invention has cytostatic activity, and may have a use
CC   in gene therapy. The set of oligonucleotides comprising at least two of
CC   the oligomers are useful for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms (SNPs) within the sequences. The
CC   methods, nucleic acid, oligonucleotide, and kit are useful for the
CC   treatment, characterisation, classification and/or differentiation, of
CC   breast cell proliferative disorders. The method is also useful for
CC   predicting the responsiveness of a subject with a cell proliferative
CC   disorder of the breast tissues to a therapy. The present sequence is used
CC   in the exemplification of the invention.
XX
SQ   Sequence 16579 BP; 4923 A; 0 C; 3197 G; 8459 T; 0 U; 0 Other;

  Query Match             4.5%;  Score 53.4;  DB 13;  Length 16579;
  Best Local Similarity   42.1%;  Pred. No. 0.083;
  Matches  303;  Conservative    0;  Mismatches  416;  Indels    0;  Gaps    0;

Qy         30 AAGGGATGTACTCAATGTATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAG 89
              |||  || || | |||   ||  | ||    ||||||  || ||  | |  |  | |||
Db      13377 AAGTTATATAATTAATAGTTTAGTGGTTTTTTTAGTGTTAATTATAGTAATTTTAGATTT 13436

Qy         90 GAAAAAAACTTATTTCCCGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATG 149
               | |      | ||||     |||      | ||   |   |||   |    |   | |||
Db      13437 AAGATTGTTTGATTTAGGAATGGAGAGAAATAAAATAAAATGAATGTTTATTGAAATATA 13496

Qy        150 GGTAATAGCGTTAACTATATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAG 209
                |  |||   | |  |||   |      |  |||  |       ||   |   |
Db      13497 TATTTTAGATATTAAGATAGATTTTTTTGTGAAATATTTTAAGATGTGGATTTGTATTTT 13556

Qy        210 TTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCA 269
              | | ||  |||   | | |||   |    ||       || ||   ||
Db      13557 TATTTTTAAATTTGAGTAAATTTAGTTGAGTTTTAAAGAGTTTAAGTAATTTGTTTAAGG 13616

Qy        270 ATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGAC 329
              | ||     | ||| | | | ||| | |  | || |||  || | |||||  ||  |
Db      13617 TTATTTTGTAAGTGATAGATTTGTTAGTTTGAAAAGTTTATTATTTTTTATTTTTTTGTT 13676

Qy        330 AATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGA 389
              | | |  ||| | |    ||   | |       || | | |  || || |  |
Db      13677 TTTTTATGATATAGTTAAGTAAAATTTTTTAATAAAGGAGAATTAGATTAGAAATAAGTA 13736

Qy        390 GCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTAC 449
                | ||   ||   |||   |||   ||  | ||  |    ||  |       |
Db      13737 ATTAATAGTAGATTAATGTTTATTTTTAGGGAAATGTTAATAGTTAATATTTATTATAAA 13796

Qy        450 AGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATG 509
                | | || ||       ||| |     | |         |||| | ||||  |
Db      13797 TAATGTTATTTGTATATAATGGAATATGATTGAAAGTGATGGGTTAGATATTTTTTATAT 13856
```

```
Qy        510 TATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCT 569
              | ||   |||||||    | | | | |||  | |||  |   | || | |  ||| |  |
Db      13857 TTTGAGATGTTTTTATTTATGTTGTTAAATTTGATTTTTTAATTTTTTAAAATTTTTAGT 13916

Qy        570 ATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTA 629
              ||| |    |    |   |  || ||  ||  ||   || ||   |     |  ||
Db      13917 TTTTTTTAATGTTAATAGATTTTATAATATTTGATTAAATATTTTTATAGGATAAAGGTA 13976

Qy        630 CGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAAT 689
              ||    |   || ||      |||   | | |  ||    ||  || |   |  |
Db      13977 TTTTTTTAGGGTTGATTTAAGGAATTAGAATATGTATAATGTTGAAAAGTTTTAAATTGT 14036

Qy        690 AATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGA 748
              | |     ||||    |||| || || |||| || | ||   || | |  |||| |
Db      14037 TTGTTTTTTTTTTTTTTTGATTTGTAATATTATTATTTAGAAAATTAAATATTTTTTTA 14095
```

RESULT 10
AAS46314
ID AAS46314 standard; DNA; 16750 BP.
XX
AC AAS46314;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #36.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 36; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 16750 BP; 4993 A; 270 C; 3224 G; 8263 T; 0 U; 0 Other;

Query Match 4.5%; Score 53.4; DB 4; Length 16750;
Best Local Similarity 42.1%; Pred. No. 0.083;
Matches 303; Conservative 0; Mismatches 416; Indels 0; Gaps 0;

```
Qy         30 AAGGGATGTACTCAATGTATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAG 89
              |||  || || | |||   ||  | ||    ||||||| || ||  | |  |  | |||
Db      13548 AAGTTATATAATTAATAGTTTAGTGGTTTTTTTAGTGTTAATTATAGTAATTTTAGATTT 13607

Qy         90 GAAAAAAACTTATTTCCCGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATG 149
              | |     | ||||     |||     | ||   |   |||   |    |   | |||
Db      13608 AAGATTGTTTGATTTAGGAATGGAGAGAAATAAAATAAAATGAATGTTTATTGAAATATA 13667

Qy        150 GGTAATAGCGTTAACTATATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAG 209
              |  |||   | |  |||   |      |  |||  |       ||   |    |
Db      13668 TATTTTAGATATTAAGATAGATTTTTTTGTGAAATATTTTAAGATGTGGATTTGTATTTT 13727

Qy        210 TTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCA 269
              |  | ||  |||   |  | |||   |    ||       || ||   ||
Db      13728 TATTTTTAAATTTGAGTAAATTTAGTTGAGTTTTAAAGAGTTTAAGTAATTTGTTTAAGG 13787

Qy        270 ATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGAC 329
              |  ||    | ||| | | | ||| | |  | || |||  || | |||||  ||  |
Db      13788 TTATTTTGTAAGTGATAGATTTGTTAGTTTGAAAAGTTTATTATTTTTTATTTTTTTGTT 13847

Qy        330 AATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGA 389
              | | |  ||| | |     ||    |  |          || | | |  || || |  |
Db      13848 TTTTTATGATATAGTTAAGTAAAATTTTTTAATAAAGGAGAATTAGATTAGAAATAAGTA 13907

Qy        390 GCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTAC 449
              | ||   ||   |||  |||  ||  | ||  |    ||  |          |
Db      13908 ATTAATAGTAGATTAATGTTTATTTTTAGGGAAATGTTAATAGTTAATATTTATTATAAA 13967

Qy        450 AGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATG 509
              | | || ||         ||| |     | |         |||| | |||| |
Db      13968 TAATGTTATTTGTATATAATGGAATATGATTGAAAGTGATGGGTTAGATATTTTTTATAT 14027

Qy        510 TATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCT 569
              | ||   |||||||    | | | | |||  | |||  |   | || | |  ||| |  |
Db      14028 TTTGAGATGTTTTTATTTATGTTGTTAAATTTGATTTTTTAATTTTTTAAAATTTTTAGT 14087

Qy        570 ATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTA 629
              ||| |    |    |   |  || ||  ||  ||   || ||   |     |  ||
Db      14088 TTTTTTTAATGTTAATAGATTTTATAATATTTGATTAAATATTTTTATAGGATAAAGGTA 14147

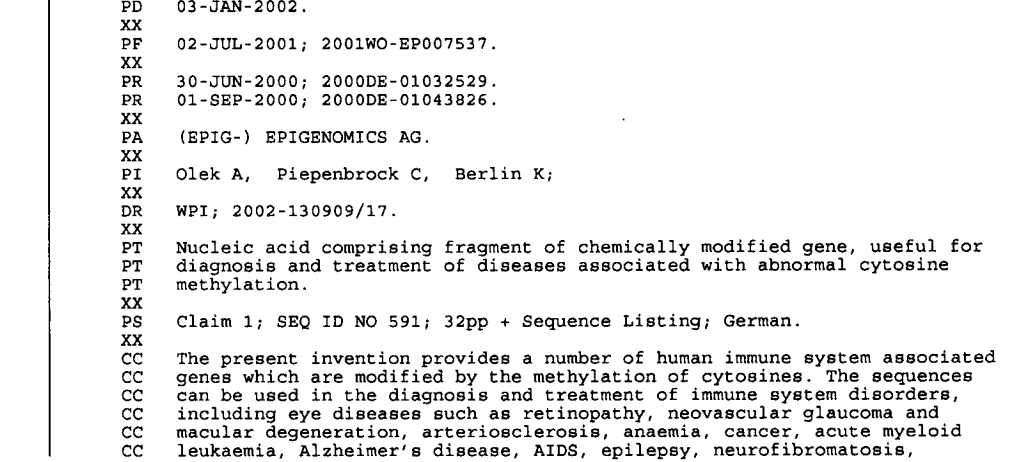
Qy        630 CGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAAT 689
              ||    |   || ||      |||   | | |  ||    ||  || |   |  |
Db      14148 TTTTTTTAGGGTTGATTTAAGGAATTAGAATATGTATAATGTTGAAAAGTTTTAAATTGT 14207

Qy        690 AATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGA 748
              | |     ||||    |||| || || |||| || | ||   || | |  |||| |
Db      14208 TTGTTTTTTTTTTTTTTTGATTTGTAATATTATTATTTAGAAAATTAAATATTTTTTTA 14266
```

RESULT 11
ABL32521
ID ABL32521 standard; DNA; 16750 BP.
XX
AC ABL32521;
XX
DT 26-MAR-2002 (first entry)
XX

DE Human immune system associated gene SEQ ID NO: 494.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 494; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 16750 BP; 4993 A; 270 C; 3224 G; 8263 T; 0 U; 0 Other;

Query Match 4.5%; Score 53.4; DB 6; Length 16750;
Best Local Similarity 42.1%; Pred. No. 0.083;
Matches 303; Conservative 0; Mismatches 416; Indels 0; Gaps 0;



RESULT 12
ABL32618
ID ABL32618 standard; DNA; 16373 BP.
XX
AC ABL32618;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 591.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 591; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

```
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 16373 BP; 4887 A; 105 C; 3275 G; 8106 T; 0 U; 0 Other;

  Query Match             4.3%;  Score 51.4;  DB 6;  Length 16373;
  Best Local Similarity   44.8%;  Pred. No. 0.23;
  Matches  246;  Conservative    0;  Mismatches  296;  Indels    7;  Gaps    1;

Qy        206 GGAGTTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACC 265
              ||| |||  | |       ||  ||  | |||  |||  | ||||| |    ||  ||
Db       8709 GGATTTAAGTGTATAAAATGGAGGTGATAATAATATTTTTTATTTTGAGTTGATGTTAGA 8768

Qy        266 CGCAATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTG 325
                | |  |||   ||||| | | || |||| ||  |||  ||| | || | ||    |
Db       8769 TTTAGTTGTTATTAATTGATAGAGTATTTTGATAATTAAAATTTTTGATTTATAAATATT 8828

Qy        326 CGACAATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATG 385
                |    |  |||| || ||       | ||   ||  |   || ||||| |   |
Db       8829 TTATTTAAGATATTTTTAATTATAAGTTTAAAATAGTTTAATTTTAATTGGTTTAAATGA 8888

Qy        386 ATGAGCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGG 445
              |  |  |  ||| |    |       ||     ||||         |  |  | | |   |||
Db       8889 AAAAATAGAATGTATTGGTTGTTATATTAGGAATTTTATAAAGTTTTTTGGTAGGGTGGG 8948

Qy        446 CTACAGGCGGTTTTA-------GTGATGCTCTCAATTTTGGATATATGCTCACATTAGGT 498
                |  |||      ||        ||  |  | | | ||||  |||  |  |    ||  ||
Db       8949 TTGTAGGTATAGTTTTATTAGGGTATTAGTTTTATTTTTTTATAGTTTTTTTAGTTTTGT 9008

Qy        499 GTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATT 558
              ||||  |||| ||| || | | |||      |||    |  |||| | |  |   | ||
Db       9009 TTTTTTTTATATATTGATTTTATTTGTAGGTAATAGGTTAAGTTATAGGTTTTATATGTT 9068

Qy        559 AGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCT 618
              | || |  |||||||||  |      |  | | ||  |  || |     || | |  |
Db       9069 TATATTATATTATTTATAGTAAGAGAGAAAATTTTTTTTTTAGAATTTTAGTAGAAGTA 9128

Qy        619 GCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACT 678
                  |       ||     |  |||     |        |  |  ||||   |||   |
Db       9129 TTAAGATGTGTTTTGATTGGATTGGTTTGTGGTATATGAGTTTTAGTTTGAAAAATTGAT 9188

Qy        679 TTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACA 738
              |||  |     | ||         ||||  |  |  ||  |      |||||  | |  ||||
Db       9189 TTAATATAGGTTATATGTTTTTTTTTTGGATATTGTTTAGCGATTTTATAGATTTATAGT 9248

Qy        739 GAAATTTTG 747
              |  |||| |
Db       9249 GGTATTTAG 9257


RESULT 13
AAD28382
ID   AAD28382 standard; DNA; 16373 BP.
XX
AC   AAD28382;
XX
DT   22-APR-2002  (first entry)
XX
DE   Human chemically treated genomic DNA #23.
XX
KW   Human; cytostatic; antidepressant; neuroleptic; nootropic; antiaddictive;
KW   adrenergic alpha-1C-receptor; cytosine methylation; therapy; alcoholism;
KW   behavioural disorder; neurological; psychiatric; cancer; schizophrenia;
KW   Tourette's syndrome; smoking; human immunodeficiency virus dementia;
KW   drug abuse; migraine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200202809-A2.
XX
PD   10-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007540.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154759/20.
XX
PT   Novel nucleic acid useful for diagnosis and therapy of behavioral
PT   disorder, neurological disorder and cancer, comprises a sequence of a
PT   segment of chemically pretreated DNA of adrenergic alpha-1C-receptor
PT   gene.
XX
PS   Claim 1; Page 114-122; 190pp; English.
XX
CC   The invention relates to nucleic acids comprising a segment of chemically
CC   pretreated DNA of adrenergic alpha-1C-receptor gene. The invention also
CC   relates to oligonucleotides or peptide nucleic acid (PNA) oligomers
CC   useful for detecting cytosine methylations. The pretreated DNA is useful
CC   for the diagnosis or therapy of behavioural disorders, neurological
CC   disorders and cancer, in particular major depressive disorder, Tourette's
CC   syndrome, schizophrenia, psychiatric and neurological disorders, smoking,
CC   drug abuse, alcoholism, personality traits, compulsive gambling, human
CC   immunodeficiency virus dementia, migraine, behaviours in schizophrenic
CC   and schizoaffective patients, and suicidal behaviour in patients with
CC   schizophrenia. The nucleic acid is useful for detecting the methylation
CC   state of all CpG dinucleotides and/or single nucleotide polymorphisms
CC   (SNPs). The present sequence is human chemically treated genomic DNA
XX
SQ   Sequence 16373 BP; 4887 A; 105 C; 3275 G; 8106 T; 0 U; 0 Other;

  Query Match             4.3%;  Score 51.4;  DB 6;  Length 16373;
  Best Local Similarity   44.8%;  Pred. No. 0.23;
  Matches  246;  Conservative    0;  Mismatches  296;  Indels    7;  Gaps    1;

Qy        206 GGAGTTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACC 265
              ||| |||  | |       ||  ||  | |||  |||  | ||||| |    ||  ||
Db       8709 GGATTTAAGTGTATAAAATGGAGGTGATAATAATATTTTTTATTTTGAGTTGATGTTAGA 8768

Qy        266 CGCAATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTG 325
                | |  |||   ||||| | | || |||| ||  |||  ||| | || | ||    |
Db       8769 TTTAGTTGTTATTAATTGATAGAGTATTTTGATAATTAAAATTTTTGATTTATAAATATT 8828

Qy        326 CGACAATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATG 385
                |    |  |||| || ||       | ||   ||  |   || ||||| |   |
Db       8829 TTATTTAAGATATTTTTAATTATAAGTTTAAAATAGTTTAATTTTAATTGGTTTAAATGA 8888

Qy        386 ATGAGCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGG 445
              |  |  |  ||| |    |       ||     ||||         |  |  | | |   |||
Db       8889 AAAAATAGAATGTATTGGTTGTTATATTAGGAATTTTATAAAGTTTTTTGGTAGGGTGGG 8948

Qy        446 CTACAGGCGGTTTTA-------GTGATGCTCTCAATTTTGGATATATGCTCACATTAGGT 498
                |  |||      ||        ||  |  | | | ||||  |||  |  |    ||  ||
Db       8949 TTGTAGGTATAGTTTTATTAGGGTATTAGTTTTATTTTTTTATAGTTTTTTTAGTTTTGT 9008

Qy        499 GTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATT 558
              ||||  |||| ||| || | | |||      |||    |  |||| | |  |   | ||
Db       9009 TTTTTTTTATATATTGATTTTATTTGTAGGTAATAGGTTAAGTTATAGGTTTTATATGTT 9068

Qy        559 AGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCT 618
              | || |  |||||||||  |      |  | | ||  |  || |     || | |  |
Db       9069 TATATTATATTATTTATAGTAAGAGAGAAAATTTTTTTTTTAGAATTTTAGTAGAAGTA 9128

Qy        619 GCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACT 678
                  |       ||     |  |||     |        |  |  ||||   |||   |
Db       9129 TTAAGATGTGTTTTGATTGGATTGGTTTGTGGTATATGAGTTTTAGTTTGAAAAATTGAT 9188

Qy        679 TTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACA 738
              |||  |     | ||         ||||  |  |  ||  |      |||||  | |  ||||
Db       9189 TTAATATAGGTTATATGTTTTTTTTTTGGATATTGTTTAGCGATTTTATAGATTTATAGT 9248
```

```
Qy        739 GAAATTTTG 747
              |  |||| |
Db       9249 GGTATTTAG 9257


RESULT 14
ACL37108/c
ID   ACL37108 standard; cDNA; 2000 BP.
XX
AC   ACL37108;
XX
DT   02-JUN-2005  (first entry)
XX
DE   Rice stress-regulated promoter SEQ ID NO:15671.
XX
KW   ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KW   agriculture.
XX
OS   Oryza sativa.
XX
PN   WO2003008540-A2.
XX
PD   30-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US019668.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   24-AUG-2001; 2001US-0314662P.
PR   26-SEP-2001; 2001US-0325277P.
PR   21-NOV-2001; 2001US-0332132P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Kreps J,  Briggs SP,  Cooper B,  Glazebrook J,  Goff SA,  Katagiri F;
PI   Moughamer T,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-248011/24.
XX
PT   New stress-responsive nucleic acid, useful for altering the
PT   responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT   stress, salt stress or osmotic stress.
XX
PS   Claim 48; SEQ ID NO 15671; 89pp; English.
XX
CC   The invention relates to novel abiotic stress responsive polynucleotides
CC   and polypeptides. Also disclosed are vectors, expression cassettes, host
CC   cells, and plants containing such polynucleotides. Also disclosed are
CC   methods for using the polynucleotides and polypeptides to alter the
CC   responsiveness of a plant to abiotic stress. The invention is useful in
CC   agriculture. The nucleic acid is useful for determining whether a test
CC   plant has been exposed to an abiotic stress condition. It is also useful
CC   for selecting an agent that alters abiotic stress regulated
CC   polynucleotide expression in a plant cell, and to identify a homolog or
CC   ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC   molecule and the polypeptide encoded by it are useful in altering the
CC   responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC   stress, osmotic stress or any of their combinations. The present sequence
CC   is used in the exemplification of the invention
XX
SQ   Sequence 2000 BP; 510 A; 350 C; 268 G; 518 T; 0 U; 354 Other;

  Query Match             4.2%;  Score 50.6;  DB 11;  Length 2000;
  Best Local Similarity   11.7%;  Pred. No. 0.22;
  Matches   55;  Conservative  213;  Mismatches  197;  Indels    4;  Gaps    1;

Qy        283 GAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATT 342
              :|  |::|   :: :    : ::::: :::: :: : : ::  |::  |:||:| ||: |
Db        482 RACCAWRTKWCKWMMRKTGKKRWWYKGMKRMARKAWGKARWAYCKRTTAYATMTTTTRRT 423

Qy        343 GATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGA 402
              :|  :::| : ::: :  ::: |:::| : ::: ::   :::| :|:   :||::   :
Db        422 RAAWWMYTGMRKRRRSMTRWRMTSRWTTWSWYKWYYYWTRRYRASRARTKGKATKYTMTW 363

Qy        403 GAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGT 462
              :::: :    :::   :    |: |: :|  : ::  ::     : :     |: |
Db        362 KWMMGCWSYTKARWWMMAYKKSWKMAWYAWTKGTRRYYYRMKRAAYWTRWMATRSATWCT 303

Qy        463 GATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTT 522
              |:|  |:    :::    :|:|| ::::  ::|||  ::::::   :|::  : : : :
Db        302 GWTYWTSAKTTKWKRWAGYAKATTMKMRKTYYAGGRWWWYKKWASSRTWKWCWYKMYMWA 243

Qy        523 TCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTAT----TTATAGC 578
              |  :::: || :|:   ::: ::  : : | |::||: :   : | :::    : | |:
Db        242 TKKSWWSWATTWAWKSMWYKKKWTYYTYMTSTYWTTWMWMRCRAGWWWKGGMAKRAAART 183

Qy        579 GATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTA 638
              ::  |  ::::  : :  :  |:|:: ::::::| :: :     : ::   ::: :|
Db        182 KWMRTYRMKKMWRYRWYKWSWARGWRAWRKRYYTYWWYMMYRCAWMWYSAYWRSWRRWTT 123

Qy        639 TATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATACC 698
              | :   ::  :: :::   : : :  ::  ::: :|::  :|:::| ::::::::   :
Db        122 TKYSKMYYWKMWMMRRTYARAKAYSRWWMWKKRCRTRWKGWTRWKTWKMMWRWYKKRTYR 63

Qy        699 CGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTG 747
              :::  :: :::   : : ::: : :   :  :: | : : : :::: ::
Db         62 MKKSAKWWYMRSGAYRRRWMRRMAYSRRRSWMRKTSYYYTKKWWKKMWS 14


RESULT 15
ABK31299
ID   ABK31299 standard; DNA; 6161 BP.
XX
AC   ABK31299;
XX
DT   23-APR-2002  (first entry)
XX
DE   Signal transduction associated gene modified complementary DNA #71.
XX
KW   Human; signal transduction associated gene; cytosine methylation state;
KW   CpG island; signal transduction associated disease; solid tumour; cancer;
KW   antitumour; cytostatic; mutant; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200200926-A2.
XX
PD   03-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007472.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-147896/19.
XX
PT   Oligonucleotide for diagnosis and therapy of diseases associated with
PT   signal transduction e.g. cancer, comprises chemically modified genomic
PT   sequences of genes associated with signal transduction.
XX
PS   Claim 1; SEQ ID NO 142; 24pp; English.
XX
CC   The present invention relates to chemically modified DNA sequences of
CC   signal transduction associated genes. The DNA sequences are chemically
CC   modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC   Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC   the cytosine methylation state (CpG islands) of these genes, and a method
CC   for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC   genes associated with signal transduction. The genomic DNA can be
CC   obtained from cells or cellular components which contain DNA, e.g. cell
CC   lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC   tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
```

CC brain, heart, prostate, lung, breast or liver, histologic object slides,
CC and all their possible combinations. The sequences of the invention are
CC useful for the diagnosis and therapy of diseases associated with signal
CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC chemically pretreated genomic DNA sequences of different genes associated
CC with signal transduction, or their complementary sequences. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office
XX
SQ Sequence 6161 BP; 1577 A; 142 C; 1452 G; 2990 T; 0 U; 0 Other;

Query Match 4.2%; Score 50.6; DB 6; Length 6161;
Best Local Similarity 43.6%; Pred. No. 0.28;
Matches 318; Conservative 0; Mismatches 409; Indels 2; Gaps 2;

```
Qy         18 AGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCGTTAGTGCTAGTTCGATT 77
              | ||  ||   |||   |  |  |    || ||  | || | || |||    ||     |
Db       2326 AAGATAAAGTTTAATTTAGTTGTTTTTGGTTTTAGTAGTTGGGTAAGTAAAGGTAAAGGT 2385

Qy         78 TGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACTATTTTTTCATGGTCGGC 137
                ||  |   ||  | || || |||    || ||||| | |    |   |||   |
Db       2386 AAAGATAAAGGCGAGTATAATTTTTTTTTTTTTTTGTTTTAATTTAAATCGGATGTCGGTT 2445

Qy        138 GGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGATAAAATTCAAGCAATTAT 197
               |   | | |   |   | | ||  |   ||   |      | |   |  |  || | |
Db       2446 AGATTTTTTTTTTTTGGGTGTGGTAGGTGGTTTACGATGTTTTGAGTGTGTATAAAGTGT 2505

Qy        198 GGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGT 257
               |||| |  | ||| | |||||||  ||  | ||||       ||  |    |    |||
Db       2506 TGGGGAGGAGGGTTTTGTTTTATTTTCGTTAATTTTTGGGAATTTTTTTTAGTATTTAGT 2565

Qy        258 ATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTT 317
               |  ||           |    |  |  ||||       |   || |   |  ||
Db       2566 TTTGTAGTTTTTGAGGAAATTATTATAGTAAATATAAGATTGTGTATTGAATTAAAGTGA 2625

Qy        318 TATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCT 377
               || |  |      || | | | ||   |    | |||  | |  | |     | |  | |
Db       2626 TAGGAATTTAGAGATTTTTTTGATAAGTTGTTTTATGGTTAGTAAATAGGGGATAGGGAT 2685

Qy        378 CGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGAT 437
                | | ||     ||||  |       |||       |  ||    |  || | |   |
Db       2686 -TTTTTTATTTTTTTTATAGAAAGATGTAATATTGTATTTTTTTAGGTAATAAAGTATT 2744

Qy        438 GGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGG 497
              |||        |     |||  ||    | |   ||||   |   |  |    | |||
Db       2745 GGTATTTTGGGAATTAAAATTTTTTGGGTGTTTTTTTTTTTTTTTCGTTTTATATTAAGG 2804

Qy        498 TGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTAT 557
              | ||| | | | | |    | |||||         | | | |  |||||  |  ||||| |
Db       2805 TATTTAGGTTTATTTTTTTTTTTTTTTTATTTTTTTTTATATTTTATTTTTTTTTATTTT 2864

Qy        558 TAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGC 617
              |||||||  ||  | |||||    |       || |   || |  ||        | || ||
Db       2865 TAGTTTTATGGTTTTTATTAGTAGGGAGTTGGGGAATAGTTAAGTTTTTTAAGTTTTTTT 2924

Qy        618 TGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAAC 677
              |  |    |||| ||  |||  | || |  || | | || |||   |       ||||||
Db       2925 TAGGGAAGTTTAGGCGATTTTGAGAAATATAAGTTGAAA-ATGTAATGGAAGGGAATAGT 2983

Qy        678 TTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATAC 737
              ||| | |  ||   |||||        |   |||| || ||    |||  | ||   |
Db       2984 TTTTTATAGAAATGTTATAGTTAAGAGATATATTTTTAATATGGATTTAAAGAATATTTT 3043

Qy        738 AGAAATTTT 746
              |||||||  |
Db       3044 AGAAATTAT 3052
```

Search completed: May 11, 2006, 09:42:56
Job time : 549.783 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 09:43:19 ; Search time 3738.81 Seconds
(without alignments)
15016.694 Million cell updates/sec

Title: US-09-423-093B-3_COPY_5114_6313
Perfect score: 1200
Sequence: 1 atgctgccatttccaccagg.....atagacggtgcacgcgataa 1200

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	68.6	5.7	1101	10	CNS0039G	AL063921 Drosophil
	2	65.8	5.5	1101	10	CNS00LT2	AL078714 Drosophil
c	3	62.6	5.2	1103	1	AJ928272	AJ928272 AJ928272
	4	60.8	5.1	2087	10	AG333887	AG333887 Mus muscu
	5	60.4	5.0	1101	10	CNS0039G	AL063921 Drosophil
c	6	59.8	5.0	1162	10	CW938442	CW938442 TcB18.3_A
c	7	58.8	4.9	1101	10	CNS00EVL	AL069706 Drosophil
	8	58.2	4.9	1101	10	CNS0100X	AL098379 Drosophil
	9	58.2	4.9	1155	1	AJ928993	AJ928993 AJ928993
	10	58.2	4.9	1259	9	CC210443	CC210443 CH261-185
	11	58	4.8	1378	10	AG350209	AG350209 Mus muscu
	12	57.6	4.8	1364	10	CG757970	CG757970 P053-2-C0
c	13	57.4	4.8	1272	1	AJ927522	AJ927522 AJ927522
	14	57.2	4.8	1161	11	CNS073Y8	AL428118 clone BA0
c	15	56.8	4.7	801	10	AG468514	AG468514 Mus muscu
c	16	55.8	4.7	1147	1	AJ925797	AJ925797 AJ925797
c	17	55.6	4.6	1101	10	CNS00Z6Z	AL097301 Drosophil
	18	55.4	4.6	1101	10	CNS00EVL	AL069706 Drosophil
	19	55.2	4.6	1177	10	AG365683	AG365683 Mus muscu
	20	54.4	4.5	941	10	CZ935951	CZ935951 255025 To
c	21	54.2	4.5	996	10	CNS00FUH	AL071063 Drosophil
c	22	53.6	4.5	1101	10	CL505531	CL505531 SAIL_751_
c	23	53.4	4.5	983	1	AJ926697	AJ926697 AJ926697
c	24	53.4	4.5	1101	10	CNS003B4	AL064082 Drosophil
	25	53.4	4.5	1246	10	AG341400	AG341400 Mus muscu
	26	53.2	4.4	1249	10	AG311105	AG311105 Mus muscu
	27	53.2	4.4	1542	10	AG386981	AG386981 Mus muscu
c	28	53	4.4	1162	1	AJ926553	AJ926553 AJ926553
c	29	52.8	4.4	977	10	CNS0140Y	AL103564 Drosophil
	30	52.6	4.4	581	11	CNS034DK	AL227297 Tetraodon
c	31	52.6	4.4	889	10	CW973509	CW973509 AIAA-aaa5
c	32	52.6	4.4	938	10	CNS006TJ	AL065906 Drosophil
c	33	52.6	4.4	1101	10	CNS00Z2U	AL097152 Drosophil
	34	52.4	4.4	987	10	CNS014PQ	AL104456 Drosophil
c	35	52.4	4.4	1101	10	CNS0021J	AL061936 Drosophil
c	36	52.2	4.4	617	9	BH182965	BH182965 022_J_15-
c	37	52.2	4.4	617	11	CNS07O8Q	AL619916 T3 end of
c	38	52.2	4.4	952	10	AG490259	AG490259 Mus muscu
c	39	52.2	4.4	1119	1	AJ926096	AJ926096 AJ926096
c	40	52.2	4.4	1225	10	CNS0161D	AL106171 Drosophil
	41	52.2	4.4	1250	10	AG390995	AG390995 Mus muscu
	42	52	4.3	736	9	CC077693	CC077693 CSU-K33r.
	43	52	4.3	816	10	CNS02KJV	AL201604 Tetraodon
c	44	52	4.3	897	10	CZ935267	CZ935267 254373 To
	45	52	4.3	1001	10	CNS0140O	AL103554 Drosophil

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

```
  Query Match             5.7%;  Score 68.6;  DB 10;  Length 1101;
  Best Local Similarity   17.2%;  Pred. No. 0.00014;
  Matches  117;  Conservative  283;  Mismatches  278;  Indels    1;  Gaps    1;

Qy          9 ATTTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCGTTAGTGCT 68
              :|::  :  ::::  :|:   :::|  | :  | ::: ::: |::| :::  ::|   |
Db       1076 WTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWWWWWWATWDTWWDKWWWWATAAKT 1017

Qy         69 AGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACTATTTTTTC 128
              :  :  : |    :: |:::|   ::::| |:|:     :  :: ::::: :: :::::
Db       1016 DTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRRDGGRKRKDKKDRKDGDDDKKG 957

Qy        129 ATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGATAAAATTCA 188
              ::::  :  ::| :: :::: :::::::  :::  ::::   : :  |: ::|:::|
Db        956 GKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDKDDDGKGKDADDDTDG 897

Qy        189 AGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGAT 248
              : :::::: :: ::  | :|:  :|:  |:::       |:::|: | |: :: ::  :
Db        896 TKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTWDAAADDWWADDRWDA 837

Qy        249 TTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTT 308
              : : ::: :|: |     : : :   ::   :: ::: :|:: ::  :  : |  :||:|
Db        836 WAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRDAADDRDDAATWTTWT 777

Qy        309 TCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGAGAGTTCAT 368
              | | :: ::: :: :|  :: |:  :  ::::: :: |||   : : ::::::::::| ::
Db        776 TTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTWKRRWKRRDTRWD 717

Qy        369 TAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTA 428
              :|:|::    ::: : |::||:  :  :  :::::::: | ::: : ::|:  ::|    |:
Db        716 DADADDTARDDRR-RRGDDGADAGKGKKTGRKRRRRDRATWDRTDAWWADAAWWTTTDTD 658

Qy        429 TGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCT 488
              |:::::::: :::  :::    |:   :::|:  ::: : : : :|: ::: | :|:|:
Db        657 TDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRADRWDRWAADTWTDAR 598

Qy        489 CACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGAT 548
              | :::|:   :   :: |   : ::: | : :||   :   | : |::| : :|||  ||
Db        597 KADRDWAKARAWRARRDRARAARADRRWTTKGKTTTATWTTWAARAAWWAWWAWATTTAT 538

Qy        549 GCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAAT 608
              ||:|| :|  ||||  |  : || |::   ::     |       : |  | |  |    |
Db        537 WTTTWTTTWTTTTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWAAAAAAAWAATTTTT 478

Qy        609 ACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTG 668
              || ||  :   ::  :|:     ||| :: | |||  : :   :|  |  |||:||
Db        477 TTTTTTTWAWTAAAWWTAWTWTTTTWTTTTWWAATTTTTTWTWTTWWAATTATTTTWTTT 418

Qy        669 TGGAATAACTTTATTTGTA 687
              :  || ||  :::::| ||
Db        417 WTAAAAAAAAWWWWWTWTA 399


RESULT 2
CNS00LT2
LOCUS       CNS00LT2               1101 bp    DNA     linear   GSS 14-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL078714
VERSION     AL078714.1  GI:5102004
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR48P19"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.5%;  Score 65.8;  DB 10;  Length 1101;
  Best Local Similarity   21.2%;  Pred. No. 0.00057;
  Matches  101;  Conservative  166;  Mismatches  210;  Indels    0;  Gaps    0;

Qy        210 TTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCA 269
              || | |||::||    : : |::    |: | :| :  |||| ||  | |  |:   :
Db        543 TTTTTTTTKWTTWKATKTDWTKKAKTAAKTAAWAAWWTTTTTTAATWTTTTKTWTTAKKT 602

Qy        270 ATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGAC 329
              :   ::|::|||| :  :  |  :||  |: ||  ||:|| : |: ||| |  ||  :
Db        603 WAWKWWAWDTATTTWTTWTTTTTWTTTTTWAATTTTTWTTTWTAWATTTTTTTTTTTAKTK 662

Qy        330 AATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGA 389
              | | | ||||| |   |  ||:  :  | | :|  ::   :     :  :|::::||
Db        663 TTTTTTTTTTATTAAATAAAAAAWTTTDTWAAAWWTTTKKKKKKAAAADKWKDAKKWDGA 722

Qy        390 GCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTAC 449
              : ||: :: :::| :| |:::::      :   : | :  : :::: :|:   |:: :
Db        723 KKATKTKKKDKKAAWAAADKKDRKKGGKKKGKGKKGKKGGKKKKKKKKGKGGWGKKGKKK 782

Qy        450 AGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATG 509
              ||: : :::| ::: :  | |   :: ::: :  |:: : | :|:  :::       |::
Db        783 AGDDDAKDKTKKKKKKAATTTTKKKKGKKKKKKAAKKKKAADRTKKTKWDAAAAAAAAKK 842

Qy        510 TATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCT 569
              | ::| ::::|:::     :: : :   : : | :: :: |: :: :: : :  |:|
Db        843 TKDKGKKKKKTKKKTTKKKKKKKKKKGGGKKDAAAKKKKKGTKKKKGKKGKGKGGGKGKG 902

Qy        570 ATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTA 629
              : :::  | ::  :: :      : : || :   :  :::|:: :: : :: : :|
Db        903 KGGKKDDDAAAAKKKGKKKKKGKGGGGKKKKGGKKGKGKGGKKKTKKTKKKKKKKKADA 962

Qy        630 CGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGT 686
              :   :: | |:||  :  :|  ::  : :: :::|:|::::|| || ::: :::::
Db        963 AKGKTKKRAAADAAAADWTDAATKKKKKKKKDKKKTKTKKKDAAAAATKKKKKKKKK 1019


RESULT 3
AJ928272/c
LOCUS       AJ928272               1103 bp    mRNA    linear   EST 10-JUN-2005
DEFINITION  AJ928272 Theileria annulata merozoite Theileria annulata cDNA clone
            tam017c06_q1k, mRNA sequence.
ACCESSION   AJ928272
VERSION     AJ928272.1  GI:67498658
KEYWORDS    EST.
SOURCE      Theileria annulata
  ORGANISM  Theileria annulata
            Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
```

```
            Theileria.
REFERENCE   1  (bases 1 to 1103)
  AUTHORS   Pain,A., Renauld,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W.,
            Kerhornou,A., Aslett,M., Bishop,R., Bouchier,C., Cochet,M.,
            Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Fosker,N.,
            Gardner,M., Goble,A., Griffiths-Jones,S., Harris,D.E., Katzer,F.,
            Larke,N., Lord,A., Maser,P., McKellar,S., Mooney,P., Morton,F.,
            Nene,V., O'Neil,S., Price,C., Quail,M.A., Rabbinowitsch,E.,
            Rawlings,N.D., Rutter,S., Saunders,D., Seeger,K., Shah,T.,
            Squares,R., Squares,S., Tivey,A., Walker,A.R., Woodward,J.,
            Dobbelaere,D.A.E., Langsley,G., Rajandream,M.-A., McKeever,D.,
            Shiels,B., Tait,A., Barrell,B. and Hall,N.
  TITLE     The genome of the host-cell transforming parasite Theileria
            annulata and a comparison with T. parva
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Pain A
            The Pathogen Sequencing Unit
            The Wellcome Trust Sanger Institute
            Genome Campus, CB10 1SA, UNITED KINGDOM
            Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of
            Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
FEATURES             Location/Qualifiers
     source          1. .1103
                     /organism="Theileria annulata"
                     /mol_type="mRNA"
                     /isolate="Ankara (clone D7)"
                     /db_xref="taxon:5874"
                     /clone="tam017c06_q1k"
                     /dev_stage="merozoite"
                     /lab_host="Bos taurus (cow)"
                     /clone_lib="Theileria annulata merozoite"
                     /note="country: Turkey:Ankara"
ORIGIN

  Query Match             5.2%;  Score 62.6;  DB 1;  Length 1103;
  Best Local Similarity   38.6%;  Pred. No. 0.0029;
  Matches  254;  Conservative    0;  Mismatches  404;  Indels    0;  Gaps    0;

Qy         39 ACTCAATGTATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAAC 98
Db       1029 ATTNATTTTTTTTTNNTATCNTTTATAAATTAATTANAATTTTTTTTATTTTTTTATATTT 970

Qy         99 TTATTTCCCGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGC 158
Db        969 TTTTTTAAAATTTTAATTTTTTAANATATTAANAAATTATNTTTTTATTAATATNTTATT 910

Qy        159 GTTAACTATATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTT 218
Db        909 CAATATTATTTTTTATNTTATTATTAATTATTTTTTTATTTTNNTNTTTATTTTTATATT 850

Qy        219 ATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAA 278
Db        849 ANNNATTNTAATTTATATATATTTNANNATATTNANAATATAATTNNTANNAAAATTNNN 790

Qy        279 TATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTAT 338
Db        789 NAATTATAATATNANTTTTTTAATATTNNTNNTNATTTATATTTTTATTTANTTNTAATT 730

Qy        339 TATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGC 398
Db        729 TATTATTNNNTANNTTATNNNANATTTTANNATTAATTTATTTTTTTATTATANNNTTAT 670

Qy        399 AGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTT 458
Db        669 AANTTATTTTTAAATATTTTNAATATNTTTTTTATTTATTTATATTTTNTTATTTTTTTT 610

Qy        459 TAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTG 518
Db        609 NTNTATTATTTTTTATTTTAATANTTAATAANATATTATTTATTTTAANNTATTTNTTTTT 550

Qy        519 TTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGC 578
Db        549 TTTTTTAAATATTNNTATATANTTATTTTTNTTTTTNTTAAATTAAATTAATATNTTTTN 490

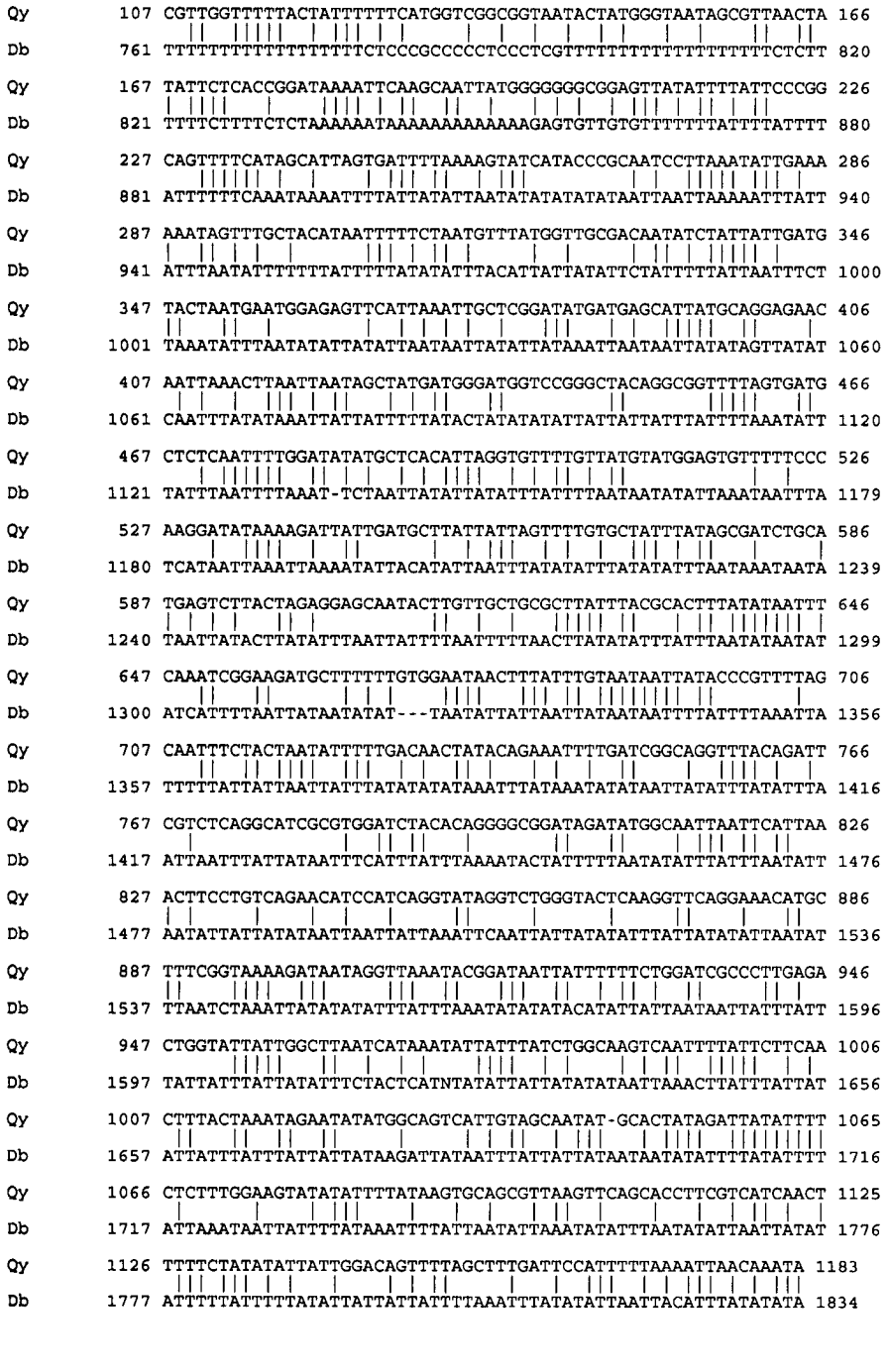
Qy        579 GATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTA 638
Db        489 ATATTTAATNATATATATTTTTNAAATTANATATATNTTTTAATAATTTTTATTATTTTA 430

Qy        639 TATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATA 696
Db        429 NATTATNTNAAATTTTAATTTNAATNNTATTATTNATATTTTTTTNATTTTNATCNTA 372


RESULT 4
AG333887
LOCUS       AG333887               2087 bp    DNA     linear   GSS 18-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-124J02.T7, genomic survey
            sequence.
ACCESSION   AG333887
VERSION     AG333887.1  GI:47907197
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 2087)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .2087
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-124J02.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             5.1%;  Score 60.8;  DB 10;  Length 2087;
  Best Local Similarity   42.5%;  Pred. No. 0.0078;
  Matches  484;  Conservative    0;  Mismatches  649;  Indels    5;  Gaps    3;

Qy         47 TATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCC 106
Db        701 TCTTTTTTTTTCTTTTTGTTTTTTTGTTTTTGTTTTTTNTGCTCCTCTGCTTTTTTTT 760
```



```
RESULT 5
CNS0039G
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.0%;  Score 60.4;  DB 10;  Length 1101;
  Best Local Similarity   18.8%;  Pred. No. 0.0091;
  Matches  119;  Conservative  250;  Mismatches  257;  Indels    8;  Gaps    1;
```



```
Qy        913 ACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGCTTAATCATAAAT 972
Db        777 AWAAWATTHHYHHTTHYMHHTYMYHYYMYTCCYMCTYHCWHYYHTAYTCWTWTHHWMWTW 836

Qy        973 ATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATATATGGCAGT 1032
Db        837 THWYHHTWWHHTTTHWAWWHTHTWCWWWWHATTWTWATHCWACMTMHWHHWMHMHHHMA 896

Qy       1033 CATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATATATTTTATAAGT 1092
Db        897 CHAHHHTHMCMCHHHMHCTCHHHHTMYHMTCHWMWHMHWHWHHWWMATWMTTMTTMMMMC 956

Qy       1093 GCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATTGGACAGTTTTA 1152
Db        957 CMMHHHCHMYHMMHMYMYCCHYYCTCHTHATTHYHYMCTCYHYCTWHTYWTAYWWAWTAH 1016

Qy       1153 GCTTTGATTCCATTTTTAAAATTAACAAATAGAC 1186
Db       1017 AMTTATWWWWMHWWAHWATWWWWWWWWATAWWAC 1050
```

RESULT 6
CW938442/c
LOCUS CW938442 1162 bp DNA linear GSS 20-DEC-2004
DEFINITION TcB18.3_A08_SP6 Tribolium BAC library Tribolium castaneum genomic, genomic survey sequence.
ACCESSION CW938442
VERSION CW938442.1 GI:56721055
KEYWORDS GSS.
SOURCE Tribolium castaneum (red flour beetle)
ORGANISM Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tribolium.
REFERENCE 1 (bases 1 to 1162)
AUTHORS Savard,J. and Tautz,D.
TITLE Tribolium castaneum BAC-ends sequencing project
JOURNAL Unpublished (2003)
COMMENT Contact: Savard, J.
Abteilung fur Evolutionsgenetik, AG Tautz
Institut fur Genetik, Universitat zu Koln
Weyertal 121, 50931 Koln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: savard@uni-koeln.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1162
/organism="Tribolium castaneum"
/mol_type="genomic DNA"
/strain="GA-2"
/db_xref="taxon:7070"
/clone_lib="Tribolium BAC library"
/note="Vector: pBACe3.6; Site_1: EcoR1; Site_2: EcoR1; Library constructed by Exelixis Inc."
ORIGIN

Query Match 5.0%; Score 59.8; DB 10; Length 1162;
Best Local Similarity 45.6%; Pred. No. 0.012;
Matches 283; Conservative 0; Mismatches 333; Indels 4; Gaps 2;

```
Qy         52 TTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCCGTTG 111
Db        797 TATATAGAAATAATAAAAATTTATATTTTTTTACATAATTTATAATTATTGTTTATATGT 738

Qy        112 GTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTC 171
Db        737 TTTTAATTAATTTTTGTTATATATTATATATTTCTATTTGTAATTATATTTAACATTTTA 678

Qy        172 TCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTT 231
Db        677 CTAATTTATATATAATTAAAATTTTATTTTATTTTAATTATTATTATTTGTTTGTAAATA 618

Qy        232 TTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATA 291
Db        617 TATTTATATTTATATTTTTTTATATTAATTTAATATTAATAATTTATTATTTTATTTATT 558

Qy        292 GTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTA 351
Db        557 ATTTTTTATATTATTTGTTTATTAGAATTTATTAATATAATAAATATAATTTAAATATTA 498

Qy        352 ATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTA 411
Db        497 TTTAT--GATGATTTAATTTATAAATTATTTATTATATTATTTATTTTTTATATTATATG 440

Qy        412 AACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTC 471
Db        439 AAATTTATAATTATATTTAATTTTAGATTTTTATTAATAAAATTTATTATTTATTTTATT 380

Qy        472 AATTTTGGATATATGCTCACATTAGGTGTTT--TGTTATGTATGGAGTGTTTTTCCCAAG 529
Db        379 ATTATTATATATATTTATATATTANTATTTTAATATTCAATATTTATATTTTTAAATTAT 320

Qy        530 GATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGA 589
Db        319 TATATAATTATAAATTATATATTTTATTATTTTTAACATTATTTATCTATTTATAAATAA 260

Qy        590 GTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAA 649
Db        259 TTTTTAATATATATATAATTATTATTACTTTATATATTATTTTTATATATTTTAATTTTA 200

Qy        650 ATCGGAAGATGCTTTTTTGT 669
Db        199 TTAATATTATAATTTATTAT 180
```

RESULT 7
CNS00EVL/c
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

```
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             4.9%;  Score 58.8;  DB 10;  Length 1101;
  Best Local Similarity   32.0%;  Pred. No. 0.02;
  Matches  180;  Conservative   90;  Mismatches  292;  Indels    0;  Gaps    0;

Qy        624 TATTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATT 683
              ||| ||    | ||:::|:||  |: |:::   :| ||  | :||| |  |||| ::|
Db       1071 TATATAAWTAATTWWTWTATATWTAWWWTATWATATATTAWTTTWTAAAATATAWWTTAA 1012

Qy        684 TGTAATAATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAAT 743
              |  |   ||:  :     | ||  :|| :  :   |:|:|::    || ::   : :  |
Db       1011 TAAATATATWTAWTAWWATATATAWATAWAAWTATAWAWTWWAATWAATWWATAWAWTTT 952

Qy        744 TTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCGCGTGGATCTACACAGGGGCGGAT 803
                | |:        :::    |||  | |       :    |  :| |  | |      |
Db        951 AATAAWAATATWAWWWTWTTATTTTTTTTTTTTTWTTTWTATWTWTTTATATWWTWWAAA 892

Qy        804 AGATATGGCAATTAATTCATTAAACTTCCTGTCAGAACATCCATCAGGTATAGGTCTGGG 863
              | |::    :  |  || :||| : ||  | | | |    |  :|    |::     :
Db        891 AWAWWAAWTWTATTTTTTWTTATWATTWATATTAWATTTTTTWTWTATTWWTATATWATW 832

Qy        864 TACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGATAATTA 923
              :  |  :  ||  :  : | :|  ||    : :       :   :|| |::   :  :::
Db        831 WTATWTWTWTTTWATWTATWTATTTATWTWTWTTWTATTWTATWTATAWWWTWWATWWWT 772

Qy        924 TTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCT 983
               :  :| |  ||     ||   : :  |:||:||   ||: |   |: |||:  |:|| :
Db        771 AWAAWTATATATWWATATTATWWTWATTWTTWTTTTATTWTTTTAAWTTATWTATWATTW 712

Qy        984 GGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAGCAA 1043
                |  |   |||| || || :  |||:  |      |  ||:   | : :|| |   |
Db        711 TTWATWTTTTTTTTTTTTTTTWTATTTWAAATTATTTAATTAWWWAATWAWTTATTTAAT 652

Qy       1044 TATGCACTATAGATTATATTTTCTCTTTGGAAGTATATATTTTATAAGTGCAGCGTTAAG 1103
               |:  | || |   |  |   | |  |:  ||     | |||||||     :   || |
Db        651 AAWTAATTAAATTATTAAAAATTTTATWAAAAAATATTTTTTTATATTAATWAWATTTAT 592

Qy       1104 TTCAGCACCTTCGTCATCAACTTTTTCTATATATTATTGGACAGTTTTAGCTTTGATTCC 1163
               :   |  |   | :| || |:||  |:||| :: ||  |   |:||   ||| |:|
Db        591 AWATTAATATAWWTTWTAAATTWTTAATWTATTWWTTTAAATTATWTTTTTTTTTAWTAW 532

Qy       1164 ATTTTTAAAATTAACAAATAGA 1185
              |::|: ||:| |:| ::::| |
Db        531 AWWTWAAAWAATWAWWWWWAAA 510


RESULT 8
CNS0100X
LOCUS       CNS0100X                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098379
VERSION     AL098379.1  GI:5609990
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN03G04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             4.9%;  Score 58.2;  DB 10;  Length 1101;
  Best Local Similarity   19.4%;  Pred. No. 0.028;
  Matches   84;  Conservative  175;  Mismatches  174;  Indels    0;  Gaps    0;

Qy        240 ATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTA 299
              |:: :|: ::|:  : :|:  |:        ::  :|: :| |||:  :: : :|::: ::
Db        427 AKWTDTRTDDTWKTDTDTWGKADTDWWTDTWKTTWTWTWTTTTGWKKWDTKTKTKWKDDD 486

Qy        300 CATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGG 359
                |: :|::: : :::|:: : : :    | ::|:: : :  ::: : :  :  |:  |::
Db        487 GTTWTDTWKDWKTDKDTDKTWTDVWADNAGWWTDKDKTKAKKKRTKTDTKKTKTRTGTKK 546

Qy        360 AGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAAT 419
                ::    |: : : :   :: :  :|::|   |:: :  : : :|   : :   |:|:|
Db        547 KTDRGMRKTDTWGWGDWRADRTDGARADRAANAKDKDADGKADTDATKTDGWGTTTKADT 606

Qy        420 TAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGG 479
              |::|:  :::  : :  : :|  |::  : :|: ||:::| :|: :: : |   |:|: |
Db        607 TDDTDADWWDWKDWDWKDADTKAGRKRGWGWGKKGGWKKTTKTKKKKTDKTKTGTDTKWG 666

Qy        480 ATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAG 539
              :::::|: : :    :|: ::: || : :::|:|:: : :  :    || :||| | ||:
Db        667 WKDKWTDKKRDGKRGDGDRKDKGTGGDGDSKAKGDRAWAWRAKATRAAAADATAAATAAK 726

Qy        540 ATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAG 599
              || : |: |  :| || ||  |||| |  | ::| ||:  :  ::   ::: ::  :::
Db        727 ATAWATKTTTKKTTTTTTTKTTTTTTTTTTTKKTTTARKADRDDDDKAKRDGDKDGDWDD 786

Qy        600 AGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGAT 659
              |:::  |  : :::| :    :    :::| : :::| ::: ::| | :| : : |   |
Db        787 ADKDWKAGRDRDDDTRKWGWKDRAGKKKKAKKKDHKTAWWWADDTGTAWATWAKTATAKT 846

Qy        660 GCTTTTTTGTGGA 672
              |  : |:::::|:
Db        847 GKAKATKDBKRGR 859


RESULT 9
AJ928993
LOCUS       AJ928993                1155 bp    mRNA    linear   EST 10-JUN-2005
DEFINITION  AJ928993 Theileria annulata merozoite Theileria annulata cDNA clone
            tam021d03_q1k, mRNA sequence.
ACCESSION   AJ928993
VERSION     AJ928993.1  GI:67499382
KEYWORDS    EST.
SOURCE      Theileria annulata
  ORGANISM  Theileria annulata
            Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
            Theileria.
REFERENCE   1  (bases 1 to 1155)
  AUTHORS   Pain,A., Renauld,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W.,
            Kerhornou,A., Aslett,M., Bishop,R., Bouchier,C., Cochet,M.,
            Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Fosker,N.,
            Gardner,M., Goble,A., Griffiths-Jones,S., Harris,D.E., Katzer,F.,
            Larke,N., Lord,A., Maser,P., McKellar,S., Mooney,P., Morton,F.,
            Nene,V., O'Neil,S., Price,C., Quail,M.A., Rabbinowitsch,E.,
            Rawlings,N.D., Rutter,S., Saunders,D., Seeger,K., Shah,T.,
```

```
            Squares,R., Squares,S., Tivey,A., Walker,A.R., Woodward,J.,
            Dobbelaere,D.A.E., Langsley,G., Rajandream,M.-A., McKeever,D.,
            Shiels,B., Tait,A., Barrell,B. and Hall,N.
  TITLE     The genome of the host-cell transforming parasite Theileria
            annulata and a comparison with T. parva
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Pain A
            The Pathogen Sequencing Unit
            The Wellcome Trust Sanger Institute
            Genome Campus, CB10 1SA, UNITED KINGDOM
            Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of
            Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
FEATURES             Location/Qualifiers
     source          1. .1155
                     /organism="Theileria annulata"
                     /mol_type="mRNA"
                     /isolate="Ankara (clone D7)"
                     /db_xref="taxon:5874"
                     /clone="tam021d03_g1k"
                     /dev_stage="merozoite"
                     /lab_host="Bos taurus (cow)"
                     /clone_lib="Theileria annulata merozoite"
                     /note="country: Turkey:Ankara"
ORIGIN

  Query Match             4.9%;  Score 58.2;  DB 1;  Length 1155;
  Best Local Similarity   26.4%;  Pred. No. 0.028;
  Matches  180;  Conservative    0;  Mismatches  502;  Indels    0;  Gaps    0;

Qy         45 TGTATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTT 104
Db        435 TTTTNTTTTTTTTTTNNTTNNNNTTNTNTTTTTTNTNNTTTNTNNNNNNTTNTNNTNNNT 494

Qy        105 CCCGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAAC 164
Db        495 TTTTTTTTTTTNNNNNNTNNNNTNNTNNTTNTNNNNNNNNNTTNNNNTTTTTNNNTNNN 554

Qy        165 TATATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCC 224
Db        555 NNTNNNTTTNNTTTTTNTNTTNNTTTNNNTTNTTTNNTNNTNNNNTTTTTTTTTNTNNNNN 614

Qy        225 GGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGA 284
Db        615 NNNNNTTTTTTTTNNTTTNNTTNTTTTTTTNNTTTTTTTTNNNNTTNNNNNNNNNTTTNTNN 674

Qy        285 AAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGA 344
Db        675 NNNNTNNNTTNNNNNTNTNNTTNTTTTNTNTTTNNNTTTNNNTTNTNTTNTTNTTNTNN 734

Qy        345 TGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGA 404
Db        735 TTTTTNNNNNNTTNNNNNTNNNNTTTNNTTTTTTTTNNTTNNTNTTNTNNNNNNNNNNTT 794

Qy        405 ACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGA 464
Db        795 TTTTNNTNNNTTNNTNNNNTTNTNNTTTTNTTTTTNNNTNNNNNTTNNNNTTTTTTTTTN 854

Qy        465 TGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTC 524
Db        855 NNNNNNNNNNTNTTTNNTNNTTTNTTTTTTTTTTTTTTNTTTTTTTTTTNTNNTNTTTTTN 914

Qy        525 CCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTG 584
Db        915 TTTTTTTTTTTTTNNNNTNTTNTTNTTNTNTNNTNTTTNNNTNTTTTTTTTTNTNTTTT 974

Qy        585 CATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAAT 644
Db        975 TNNNNTTTNTTTNNNNNNNTTTTTTTTTTTTTTNTTTNNTNTTTNNTTTTTTTTTTTNNNT 1034

Qy        645 TTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTT 704
Db       1035 TTNNNTTTTTTNNNTTTTTTTTTTTNNNTTTTTTTNTTTNNTTNNNTTTTTNNNTNNTNN 1094

Qy        705 AGCAATTTCTACTAATATTTTT 726
Db       1095 NNTTTTNTNNTTTTTTNTTNTT 1116


RESULT 10
CC210443
LOCUS       CC210443               1259 bp    DNA     linear   GSS 12-MAY-2003
DEFINITION  CH261-185J8_RM1.2 CH261 Gallus gallus genomic clone CH261-185J8,
            genomic survey sequence.
ACCESSION   CC210443
VERSION     CC210443.1  GI:30529111
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1259)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: RM1 TACGACTCACTATAGGGAGA
            Class: BAC ends
            High quality sequence start: 24
            High quality sequence stop: 688.
FEATURES             Location/Qualifiers
     source          1. .1259
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-185J8"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN

  Query Match             4.9%;  Score 58.2;  DB 9;  Length 1259;
  Best Local Similarity   45.0%;  Pred. No. 0.028;
  Matches  156;  Conservative    0;  Mismatches  191;  Indels    0;  Gaps    0;

Qy        392 ATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAG 451
Db        639 ATTTTGATGGAAAGTGTTGCATTTTAAAGAAATGGAGTGGTTGGGGGGCGGNNNNNNNNN 698

Qy        452 GCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTA 511
Db        699 NNNNATTNNNTNNNTNNNTTTNTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 758

Qy        512 TGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTAT 571
Db        759 TTTTTTTTTTTTTATTTTTTTTTTTNTNTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTT 818

Qy        572 TTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACG 631
Db        819 TTTTTTTTTTTTTTTTTTTTTTTTTTTTATATTAATNATTATTATATTATTTAAATCGT 878

Qy        632 CACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAA 691
Db        879 ATTATAATATTTTATAAATATTAATTATTAATTTTTTTAAATTAATATAATTATAATTTA 938

Qy        692 TTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACA 738
```

```
Db        939 TTATTAATTTTTTATTAATTTTATATATTATTTATCATAATTATATA 985


RESULT 11
AG350209
LOCUS       AG350209               1378 bp    DNA     linear   GSS 18-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-146M02.TJ, genomic survey
            sequence.
ACCESSION   AG350209
VERSION     AG350209.1  GI:47923519
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1378)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1378
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-146M02.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             4.8%;  Score 58;  DB 10;  Length 1378;
  Best Local Similarity   43.5%;  Pred. No. 0.031;
  Matches  420;  Conservative    0;  Mismatches  535;  Indels   11;  Gaps    3;

Qy        209 GTTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGC 268
              ||||||   ||||       |||||  ||  |||  |   |||  |||   |  ||
Db        234 GTTATAGAATATTGTGATATGTTTTTTTAATATTTATTTATTTGAAATATTGTTAGTATT 293

Qy        269 AATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTT---ATGGTTG 325
              |    |  | | ||  | || ||  ||  ||  ||  | ||| | || | | |    |  |
Db        294 ATATATATATTGTTATATAATTATATTATTATTTTTTTTATATAGTTTATTGATTTTATT 353

Qy        326 CGACAATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATG 385
              ||   | |    ||||  | |   | |   || | |  |  ||   ||  |   ||||
Db        354 TGATTTTTTAATGTATTATTTTTTAATTATTTGAAAATATTTTTTTTTTTTTATAATATT 413

Qy        386 ATGAGCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGG 445
              ||||  |  | |   || | | ||| |    |   |||  |       |
Db        414 TATTTTTTTATTTATTATATTTATAATAATTACTCCTTTTTTATTTTATTTAATTTTTAT 473

Qy        446 CTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGT 505
              ||      ||||  | ||    |  || | |||||  |   ||  | | ||  |
Db        474 TTATTTATTTTTTTTTTTATTTATTTTTTTATTTTTATATTTTTTTTTTTTTTTATTAT 533

Qy        506 TATGTATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTG 565
              | |||  | | | |||     |   | || |   |||||     || || ||| ||||
Db        534 GTTTTATTTATTATATTTAATTAATTTTTATATATTTATTATATATTTTTTTTATTTTTA 593

Qy        566 TGCTATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTA 625
              |  | |||  |    || |  || |   ||  |  |  |  | |  || ||  |     |
Db        594 TAATTTTTTTTATTAT-TATATTATAATTTTTTTATTATTATTTTTTATTTTTTTTTATT 652

Qy        626 TTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTG 685
              ||||  | ||| | |  ||  | |     |  |||||  |  |  || | | |||
Db        653 TTTATTAATTTTTTTTTTTTATTATTTATTTATTATTTTTTATTTTTTTATATATTTTTT 712

Qy        686 TAATAATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTT 745
              ||  ||||     |  |     ||| |  |||| ||  |     |||     | | |
Db        713 ATATTTTTATTTATTTAATTATTTTTTTTTATAATTTTAATTTATTTTATTTTTTATTAT 772

Qy        746 TGATCGGCAGGTTTACAGATTCGTCTCAGGCATCGCGTGGATCTACACAGGGGCGGATAG 805
              |          |||      |  | |       |   | | |  |         |||
Db        773 TATATTTTTTTTTTTATTTTATTTTTATTTATTTATATTATTATTAATTTTTTTATATAT 832

Qy        806 ATATGGCAATTAATTCATTAAACTTCCTGTCAGAACATCCATCAGGTATAGGTCTGGGTA 865
              || |    |||| ||  | |  ||  | |  |  ||     |  | ||  | |  |
Db        833 ATTTTTATATTATTTTTATTATTTTTTTTTTTTTATAATTATATATTTTTATTTTTTTATT 892

Qy        866 CTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGATAATTATT 925
              |     ||      |   |  ||  ||  |   |  |   ||        || ||||||
Db        893 TTTTTTTTTATTATATTTTAAATTATGTTTATTTTTTTTTTTT-------TATTATTATT 945

Qy        926 TTTTCTGGATCGCCCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCTGG 985
              || | |    |      || | | |  |||||||   ||| |  |   |  | |||   |
Db        946 TTATTTTTGTTTTTATTTATATTTTTATTATTTTTTTTATTATTTTTTTATTTTTTATTTT 1005

Qy        986 CAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAGCAATA 1045
              |   | |||  ||| ||  | |||| |  || | | |  | | ||  |  |  | |
Db       1006 TTATATATTTTAATTATTTTAATTTAATTTTTAAATTTTTTTATATTTTTTTTTTTTAAA 1065

Qy       1046 TGCACTATAGATTATATTTTCTCTTTGGAAGTATATATTTTATAAGTGCAGCGTTAAGTT 1105
              |    | ||  ||   || | | | |     | | |||| ||| | |      ||   ||
Db       1066 TTATATTTAATTTTATTTATATGTATTTTTATTTTTATTATATTATTATTTTTTTTTATT 1125

Qy       1106 CAGCACCTTCGTCATCAACTTTTTCTATATATTATTGGACAGTTTTAGCTTTGATTCCAT 1165
              ||  | ||   | ||||  |||||  | |||      |||||   ||  |||  ||
Db       1126 TTTATATTTTTTTATATATTTTTATTATATTTAATTTTTATTTTTTAATATTTATTATAT 1185

Qy       1166 TTTTAA 1171
              ||| |
Db       1186 ATTTTA 1191


RESULT 12
CG757970
LOCUS       CG757970               1364 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P053-2-C06.zb Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG757970
VERSION     CG757970.1  GI:37987066
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1364)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
```

```
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1364
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN

  Query Match             4.8%;  Score 57.6;  DB 10;  Length 1364;
  Best Local Similarity   45.0%;  Pred. No. 0.038;
  Matches  249;  Conservative    0;  Mismatches  299;  Indels    5;  Gaps    1;

Qy        635 TTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTA 694
              | || ||| ||| | ||   |   | |||||| |   || | | || || | ||| |||
Db        560 TATAAATATTTTTATATATAATTTTTTTTTTTTTTTTTATTATTATAATTTTTATATTTA 619

Qy        695 TACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATT-----TTGAT 749
              ||    |||||    ||  || ||    ||      |  || | | |||     || ||
Db        620 TATTTTTTTTATTTTTTATTATTATATATTAAATTATATAAATAAATATTAAATATTTAT 679

Qy        750 CGGCAGGTTTACAGATTCGTCTCAGGCATCGCGTGGATCTACACAGGGGCGGATAGATAT 809
              |  | |    ||   |   |           | |          |||  |||
Db        680 TTTAAATTATTATTATNATTTATAATTTNTATAATTTTTTTTTTTAATTATAATATTTAT 739

Qy        810 GGCAATTAATTCATTAAACTTCCTGTCAGAACATCCATCAGGTATAGGTCTGGGTACTCA 869
              | ||  ||  | ||| ||  | | |  |  |  |  |    |   |    |   |
Db        740 TTTATTTTTTTATTAAAATTTTTTATAATTAATTTAAATATTATTTATTTATTTTTAAA 799

Qy        870 AGGTTCAGGAAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGATAATTATTTTTT 929
              |  || |   || ||  |||    ||||  | |||   || |     |  |||  | |
Db        800 AAATTTATTTAANATTTTTTATTAAAAATTTTATAATATATAATTAAAATATTTATATAA 859

Qy        930 CTGGATCGCCCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCTGGCAAG 989
              |  |          | | |  |  ||       |||| ||| | |  | |||   |   ||
Db        860 NTATAATTAATAAAATAATTTTTATAAATAAATAATTATATAAAAAAATTAATTTATAAT 919

Qy        990 TCAATTTTATTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAGCAATATGCA 1049
              |  ||   ||   | || ||||  ||||| | ||||    |    ||  ||  |      |
Db        920 TATATAAAATAAATTAATTTTAAAAAATAAATTATAATTAAAAATTTTTAATATATAAAA 979

Qy       1050 CTATAGATTATATTTTCTCTTTGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGC 1109
                ||| ||| | |  | | ||   || ||||||||    |    |   ||| || |
Db        980 AAATATATTTTTTAATTTATTATAAATTATATATTAATAATTATAATTTATAATTTTATT 1039

Qy       1110 ACCTTCGTCATCAACTTTTTCTATATATTATTGGACAGTTTTAGCTTTGATTCCATTTTT 1169
              |  ||  |  | ||   ||  | ||| |||||    |  | ||  | | |    || | |
Db       1040 ATATTTATATTTAAANATTAATTTATTTTATTTATTATATATATTTATTAAATAATATAT 1099

Qy       1170 AAAATTAACAAAT 1182
                | || ||  |||
Db       1100 TATATAAATTAAT 1112


RESULT 13
AJ927522/c
LOCUS       AJ927522               1272 bp    mRNA    linear   EST 10-JUN-2005
DEFINITION  AJ927522 Theileria annulata merozoite Theileria annulata cDNA clone
            tam007c01_q1k, mRNA sequence.
ACCESSION   AJ927522
VERSION     AJ927522.1  GI:67497906
KEYWORDS    EST.
SOURCE      Theileria annulata
  ORGANISM  Theileria annulata
            Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
            Theileria.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Pain,A., Renauld,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W.,
            Kerhornou,A., Aslett,M., Bishop,R., Bouchier,C., Cochet,M.,
            Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Fosker,N.,
            Gardner,M., Goble,A., Griffiths-Jones,S., Harris,D.E., Katzer,F.,
            Larke,N., Lord,A., Maser,P., McKellar,S., Mooney,P., Morton,F.,
            Nene,V., O'Neil,S., Price,C., Quail,M.A., Rabbinowitsch,E.,
            Rawlings,N.D., Rutter,S., Saunders,D., Seeger,K., Shah,T.,
            Squares,R., Squares,S., Tivey,A., Walker,A.R., Woodward,J.,
            Dobbelaere,D.A.E., Langsley,G., Rajandream,M.-A., McKeever,D.,
            Shiels,B., Tait,A., Barrell,B. and Hall,N.
  TITLE     The genome of the host-cell transforming parasite Theileria
            annulata and a comparison with T. parva
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Pain A
            The Pathogen Sequencing Unit
            The Wellcome Trust Sanger Institute
            Genome Campus, CB10 1SA, UNITED KINGDOM
            Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of
            Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="Theileria annulata"
                     /mol_type="mRNA"
                     /isolate="Ankara (clone D7)"
                     /db_xref="taxon:5874"
                     /clone="tam007c01_q1k"
                     /dev_stage="merozoite"
                     /lab_host="Bos taurus (cow)"
                     /clone_lib="Theileria annulata merozoite"
                     /note="country: Turkey:Ankara"
ORIGIN

  Query Match             4.8%;  Score 57.4;  DB 1;  Length 1272;
  Best Local Similarity   43.1%;  Pred. No. 0.042;
  Matches  382;  Conservative    0;  Mismatches  494;  Indels   10;  Gaps    2;

Qy        293 TTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAA 352
              |||  |   |  ||||  | || | | |  ||   |  |||| |  | || || || | |
Db       1270 TTTTTTTTTTTTTTTTATTTATTTATTTATTTAATATTATATGTTATTTTAATTTAATTA 1211

Qy        353 TGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAA 412
              ||  |      || ||||| ||  |   ||          ||     | |    | ||
Db       1210 TGTTTTTTATATTTATTAATTTATTTTTATTATTAATTTTTTTATATGTAATTTATTTTT 1151

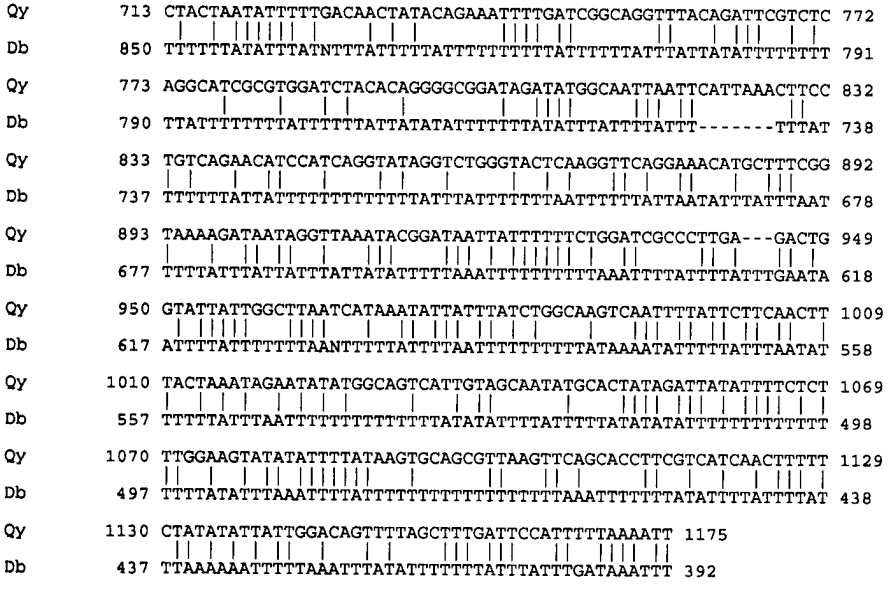
Qy        413 ACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCA 472
              |  || ||| |   ||   |   |          |          || |  |  |  | | |
Db       1150 ATATATTTATTTTTTANATTTATTTAAATTAATTTTTTATTATTTATTTATTTATTAT 1091

Qy        473 ATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGAT 532
              | ||  ||||      |  |    | | ||   || |||    | ||| |    |  |
Db       1090 TTATTTTATATTATTATATTTATTTTTTATTTAAATATATATTATATTTATATTTATTAA 1031

Qy        533 ATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTC 592
              ||   | || |  ||  |||||||||| | || |  | ||  |        |   |  |
Db       1030 ATTTTATTTTTTATATTTTTATTATTATTATTATTTTTTTATTTTTATATTATTTATTTT 971

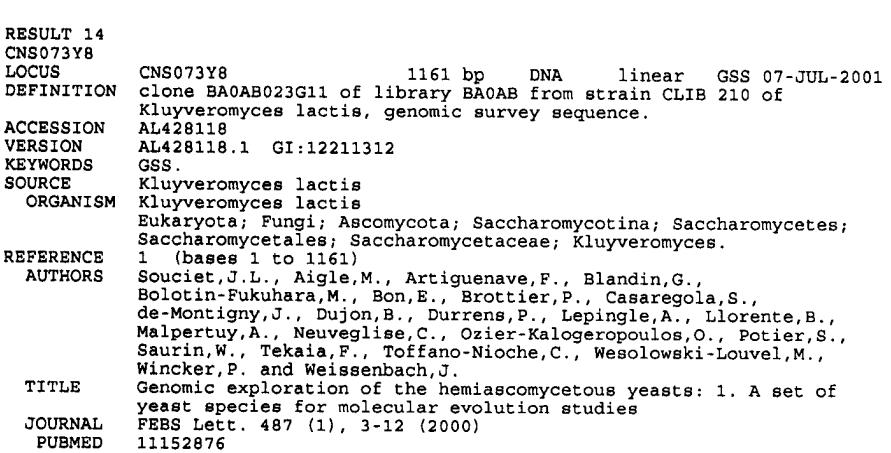
Qy        593 TTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATC 652
              || ||   |  | |  || ||  |   | |||||      ||||| |||| |  |
Db        970 ATATTATTTTATTATTTTTTTTTTTTTTTTTTAATTTATATTTATTATATTATTTTATTTT 911

Qy        653 GGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTT 712
                  ||| | |        |||||||   ||  || |     | |      | |
Db        910 TTTTTTATTATTTATTTTTTTATTTTTTATTTTATATTTTTTTTATTTTATATTTTAATA 851
```



RESULT 14
CNS073Y8
LOCUS CNS073Y8 1161 bp DNA linear GSS 07-JUL-2001
DEFINITION clone BA0AB023G11 of library BA0AB from strain CLIB 210 of Kluyveromyces lactis, genomic survey sequence.
ACCESSION AL428118
VERSION AL428118.1 GI:12211312
KEYWORDS GSS.
SOURCE Kluyveromyces lactis
ORGANISM Kluyveromyces lactis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE 1 (bases 1 to 1161)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
PUBMED 11152876
REFERENCE 2 (bases 1 to 1161)
AUTHORS Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F., Duchateau-Nguyen,G., Lemaire,M., Marmeisse,R., Montrocher,R., Robert,C., Termier,M., Wincker,P. and Wesolowski-Louvel,M.
TITLE Genomic exploration of the hemiascomycetous yeasts: 11. Kluyveromyces lactis
JOURNAL FEBS Lett. 487 (1), 66-70 (2000)
PUBMED 11152886
REFERENCE 3 (bases 1 to 1161)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES Location/Qualifiers
source 1. .1161
/organism="Kluyveromyces lactis"
/mol_type="genomic DNA"
/strain="CLIB 210"
/variety="lactis"
/db_xref="taxon:28985"
/clone="BA0AB023G11"
/clone_lib="BA0AB"
ORIGIN

Query Match 4.8%; Score 57.2; DB 11; Length 1161;
Best Local Similarity 41.2%; Pred. No. 0.047;
Matches 206; Conservative 31; Mismatches 261; Indels 2; Gaps 1;



RESULT 15
AG468514/c
LOCUS AG468514 801 bp DNA linear GSS 22-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-358P20.TJ, genomic survey sequence.
ACCESSION AG468514
VERSION AG468514.1 GI:48165771
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1

```
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 801)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .801
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-358P20.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             4.7%;  Score 56.8;  DB 10;  Length 801;
  Best Local Similarity   43.5%;  Pred. No. 0.056;
  Matches  208;  Conservative    0;  Mismatches  270;  Indels    0;  Gaps    0;

Qy        249 TTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTT 308
              | || || |   |||     |||   | || |   |  ||||| | |  || |||||  |
Db        736 TATATAATTTAAATAAATATAATATATTAAAAAAAATTAAATATTATTATATATAATAAT 677

Qy        309 TCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGAGAGTTCAT 368
                || |  | ||  ||   | ||||| ||||| |      |   |  |||  |   || ||
Db        676 AATATTAATAATATTTTATAAAATATATATTAATTTATAATATAATAATATATTATTAAT 617

Qy        369 TAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTA 428
              | ||   |     |  || |  |||||  |  |  | || |||| ||||   | |  ||
Db        616 AATATATATAAATAAATATTAATATTATTAATAAATATAAATAAATTTAAAATAAAAATA 557

Qy        429 TGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCT 488
                ||  |       |       | ||    ||  | |   ||||   | | |  |
Db        556 AATAAATATAATATATTAAATTATAAATATTTTATATATTTTTTTTTTTTTTTTTTTAT 497

Qy        489 CACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGAT 548
                ||   | |||| || | | |    | |||||        | |      || ||  |
Db        496 TTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTNNTTTTTTTTTNTTTTTTTTTT 437

Qy        549 GCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAAT 608
              || || ||  |||| |  | ||| |     | |   |   | ||  |          |
Db        436 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTT 377

Qy        609 ACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTG 668
              || ||  |    || |||      ||| | |  |||    |          |||| |
Db        376 TTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTNTTTTTATT 317

Qy        669 TGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAATATTTTT 726
              |    |   ||| |||    | ||       ||||     | | |  |  | | |||
Db        316 TTTTTTTTTTTTTTTTTNTTTNNTTNNTTTTTTTTTTTTTNTTNTNTTTTTNTNTNTTT 259

Search completed: May 12, 2006, 03:30:18
Job time : 3751.81 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 19:15:19 ; Search time 164.123 Seconds
(without alignments)
12996.800 Million cell updates/sec

Title: US-09-423-093B-3_COPY_5114_6313
Perfect score: 1200
Sequence: 1 atgctgccatttccaccagg.....atagacggtgcacgcgataa 1200

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	63.8	5.3	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	2	53.2	4.4	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	3	51	4.2	74790	3	US-09-949-016-15321	Sequence 15321, A
c	4	46.4	3.9	55886	3	US-09-949-016-15129	Sequence 15129, A
c	5	46.2	3.9	601	3	US-09-949-016-132805	Sequence 132805,
c	6	46	3.8	601	3	US-09-949-002-7549	Sequence 7549, Ap
	7	46	3.8	25590	3	US-09-949-002-777	Sequence 777, App
c	8	46	3.8	187169	3	US-09-949-016-12776	Sequence 12776, A
c	9	46	3.8	191569	3	US-09-949-016-15940	Sequence 15940, A
c	10	45.6	3.8	601	3	US-09-949-002-7550	Sequence 7550, Ap
	11	45.2	3.8	50383	3	US-09-949-016-17600	Sequence 17600, A
c	12	45	3.8	601	3	US-09-949-016-132806	Sequence 132806,
	13	45	3.8	194714	3	US-09-949-016-11869	Sequence 11869, A
	14	45	3.8	196714	3	US-09-949-016-15474	Sequence 15474, A
c	15	44.6	3.7	700	3	US-09-735-271-1038	Sequence 1038, Ap
c	16	44.4	3.7	1368	3	US-09-248-796A-11323	Sequence 11323, A
	17	44.2	3.7	38371	3	US-09-949-016-12061	Sequence 12061, A
	18	44.2	3.7	38371	3	US-09-949-016-12488	Sequence 12488, A
	19	44.2	3.7	38371	3	US-09-949-016-15596	Sequence 15596, A
	20	44.2	3.7	38371	3	US-09-949-016-15597	Sequence 15597, A
c	21	43.2	3.6	27250	3	US-09-949-016-15537	Sequence 15537, A
c	22	43.2	3.6	319608	3	US-09-539-333D-1	Sequence 1, Appli
c	23	43.2	3.6	319608	3	US-09-679-409-1	Sequence 1, Appli
c	24	43	3.6	612	3	US-09-902-540-1357	Sequence 1357, Ap
	25	42.8	3.6	832	3	US-09-621-976-2813	Sequence 2813, Ap
c	26	42.8	3.6	972	3	US-09-270-767-6833	Sequence 6833, Ap
c	27	42.8	3.6	972	3	US-09-270-767-22115	Sequence 22115, A
c	28	42.8	3.6	1055	3	US-09-806-708B-23	Sequence 23, Appl
c	29	42.6	3.6	601	3	US-09-949-002-7548	Sequence 7548, Ap
c	30	42.6	3.6	18773	3	US-09-949-016-14164	Sequence 14164, A
	31	42.2	3.5	1159	3	US-09-270-767-15130	Sequence 15130, A
	32	42.2	3.5	4648	3	US-09-207-914-24	Sequence 24, Appl
	33	42.2	3.5	4655	3	US-09-207-914-20	Sequence 20, Appl
c	34	42	3.5	2559	3	US-09-569-098A-109	Sequence 109, App
	35	41.8	3.5	1055	3	US-09-806-708B-23	Sequence 23, Appl
	36	41.6	3.5	22786	3	US-09-949-016-13399	Sequence 13399, A
	37	41.2	3.4	601	3	US-09-949-016-156535	Sequence 156535,
c	38	41.2	3.4	4920	3	US-09-269-874A-1	Sequence 1, Appli
	39	41.2	3.4	32392	3	US-09-662-254B-27	Sequence 27, Appl
	40	41.2	3.4	50000	3	US-09-662-254B-24	Sequence 24, Appl
	41	41.2	3.4	51273	3	US-09-949-016-13018	Sequence 13018, A
	42	41.2	3.4	59519	3	US-09-949-016-13504	Sequence 13504, A
c	43	41.2	3.4	231129	3	US-09-949-016-16110	Sequence 16110, A
c	44	41.2	3.4	266293	3	US-09-949-016-11934	Sequence 11934, A
	45	41	3.4	9447	3	US-09-949-016-14649	Sequence 14649, A

ALIGNMENTS

RESULT 1
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 5.3%; Score 63.8; DB 3; Length 1141;
Best Local Similarity 12.1%; Pred. No. 1.9e-06;
Matches 110; Conservative 335; Mismatches 457; Indels 6; Gaps 3;

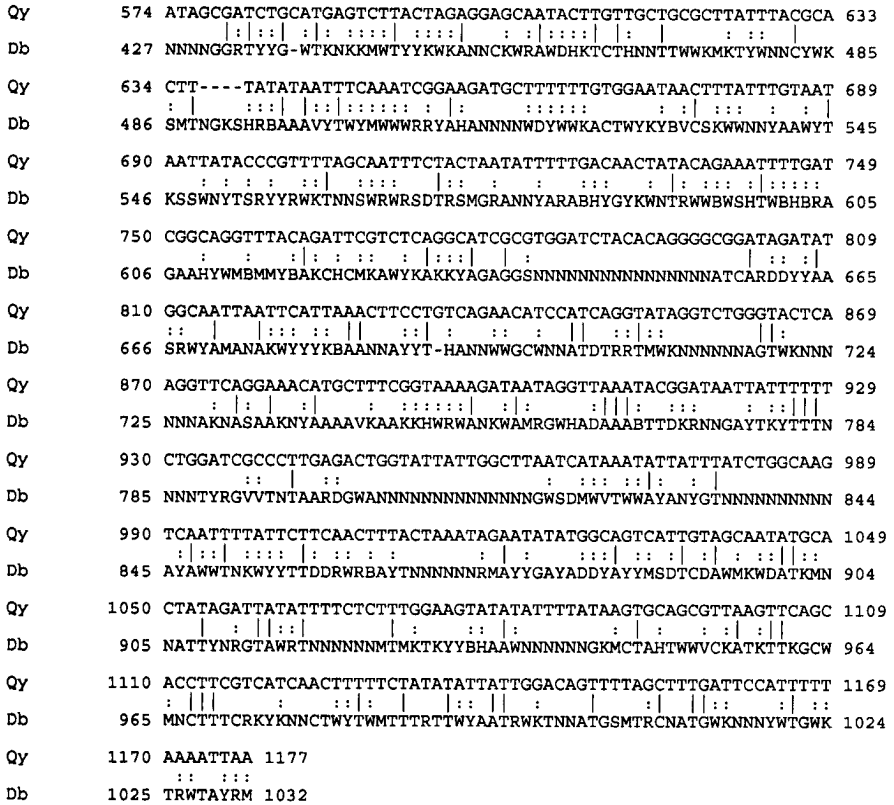
```
Qy      274 TTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATA 333
            ::::: ::: :::: :||      |:|: : ::: | :::::   ||| ::: : :::
Db      127 YWRWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNATGWDDDTKYHMWNN 186

Qy      334 TCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCAT 393
                |  : ::::|: :: ::::  |::: :   :::  :: :: :  ::::: :: :
Db      187 NGCBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRKVRRWVRTRG 246

Qy      394 TATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGC 453
              :  ::: :: |::: :    |:|: ::::          :   :: :: |      : : |
Db      247 RMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNNAKAMCKRAKYWGWNRABVNSTC 306

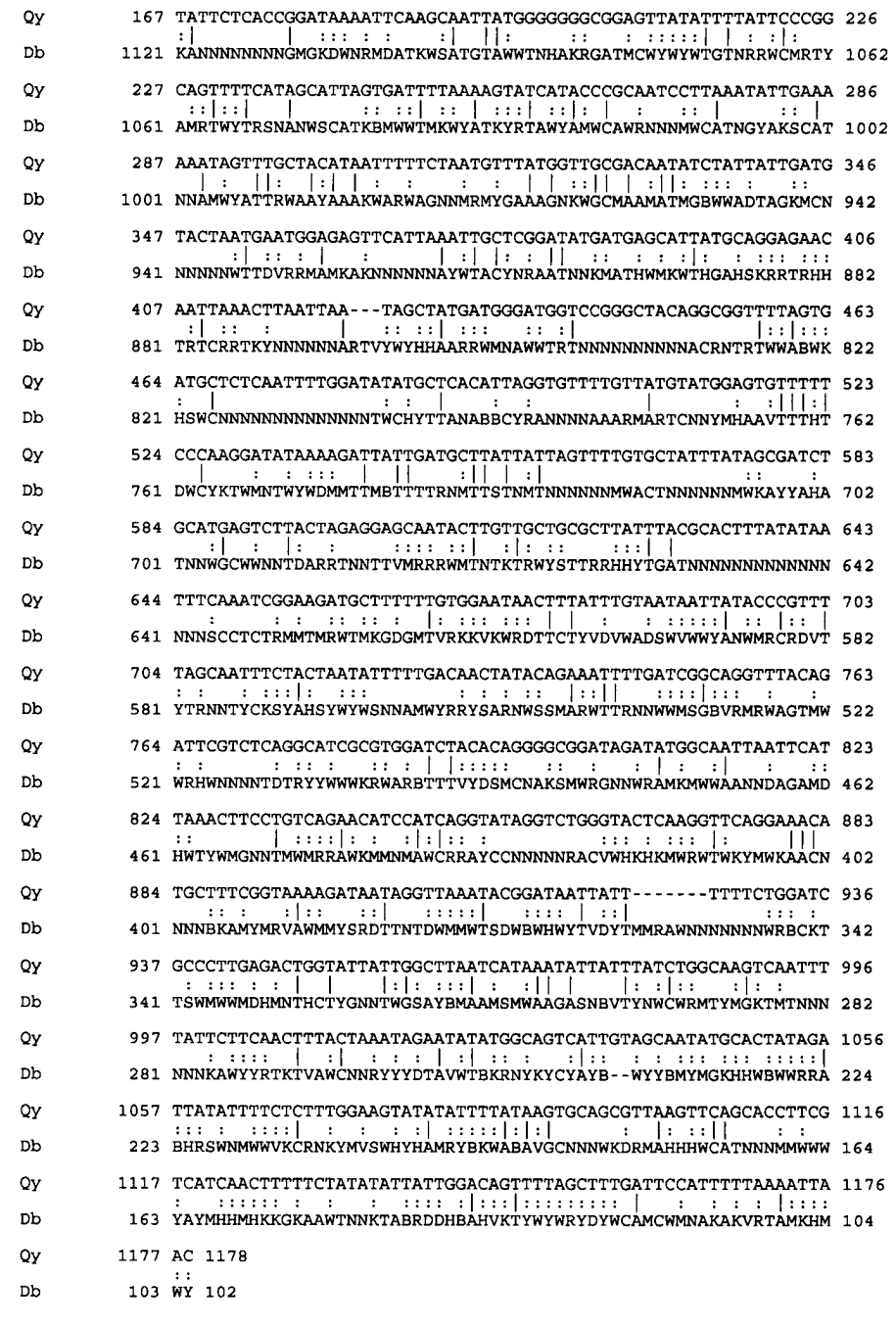
Qy      454 GGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATG 513
              :: :  :::|:|:    :  :  ::: ::: :  : :::       :| ::|  :|
Db      307 TTWKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNNWTYKKARHBARW 366

Qy      514 GAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTT 573
            ::::   ::: :  |    :: :     :          ||:::: :| ::: : : : |:
Db      367 DWVWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYN 426
```



RESULT 2
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 4.4%; Score 53.2; DB 3; Length 1141;
Best Local Similarity 12.0%; Pred. No. 0.0009;
Matches 123; Conservative 362; Mismatches 525; Indels 12; Gaps 3;



```
RESULT 3
US-09-949-016-15321/c
; Sequence 15321, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15321
;   LENGTH: 74790
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15321

  Query Match             4.2%;  Score 51;  DB 3;  Length 74790;
  Best Local Similarity   52.1%;  Pred. No. 0.013;
  Matches  114;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        958 GGCTTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAAT 1017
              ||| | ||   ||| |  | ||  | |||      ||||||||  ||| | || | |  |
Db      65737 GGCATTATTTAAAAGAGAAGTTGCCAGGCTTTCAAATTTTATAGTTCCAGTTAAATGTAA 65678

Qy       1018 AGAATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGT 1077
              |  ||| | | ||  |||| |  || ||    |   ||||  | | |||   | |     |
Db      65677 AAGATAGAAGCCAAGCATTTTTACACTAATAAAAGTAGAGAAGAATTTTAATATAATCCT 65618

Qy       1078 ATATATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATAT 1137
              || |||   ||| |  ||| |||||  ||       |  ||| | |||| | |   ||
Db      65617 TTAGATTAGCTAAATTTAGCTTTAAGAACAATTTGCACTACATTACCTTTATTTTCCTAA 65558

Qy       1138 TATTGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTA 1176
              | ||   || |||    |  |     || || |   |||
Db      65557 TTTTTTGCATTTTACATTGCGTGAGGATGTTAATTTTTA 65519


RESULT 4
US-09-949-016-15129/c
; Sequence 15129, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15129
;   LENGTH: 55886
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(55886)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15129

  Query Match             3.9%;  Score 46.4;  DB 3;  Length 55886;
  Best Local Similarity   50.4%;  Pred. No. 0.18;
  Matches  113;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        962 TAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAA 1021
              ||||| ||  |  | | ||| |   |  |  |   | || || |  |||  || ||| |
Db       5408 TAATCTTATTTTATTTATATTTCAAATATATAAAATCTTATTTACATTTCATATATATAT 5349

Qy       1022 TATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATAT 1081
               | ||  | | |   |    ||||  |  | |  |||| || | | |||  |  |||||
Db       5348 AAAATCTTATTTACATTTCATATATATATAAAATCTTATTTTATATATTTTTATATATAT 5289

Qy       1082 ATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATT 1141
              || ||||  |  |   |||  ||    |  ||    ||  | ||| | ||||||  ||
Db       5288 TTTATATATATAAAATCTTATTTTATATATATTTTATATATATTTTATATATATAAAATC 5229

Qy       1142 GGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGA 1185
                |  | |      |  | ||  |||||  |  ||    ||| |
Db       5228 ATATTTTATATATATAAAATCTTATTTTATATATATTTTATATA 5185


RESULT 5
US-09-949-016-132805/c
; Sequence 132805, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 132805
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-132805

  Query Match             3.9%;  Score 46.2;  DB 3;  Length 601;
  Best Local Similarity   48.9%;  Pred. No. 0.043;
  Matches  149;  Conservative    1;  Mismatches  154;  Indels    1;  Gaps    1;

Qy        879 AAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGATAATTATTTTTTCTGGATCGC 938
              ||| |   | | |  |||||||| ||  |  | |    |||  |||||| | |   |
Db        465 AAATAGTATCTTGAGAAAAGATACTATATATATTTATTATATATATTTTATATATTTATT 406

Qy        939 CCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCTGGCAAGTCAATTTTA 998
                |   |    ||||||       ||     ||||  |||||  |   |  |  | | ||
Db        405 ATATATAATATATATTATATATAATATATATTATATATTTTATATATTATATATAATATA 346

Qy        999 TTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAG-CAATATGCACTATAGAT 1057
              |  || |  | || || |||  ||||||   |    || ||   :||||  | |||| ||
Db        345 TATTTTATATATAATATATAATATATATTATATATTTTATATATWATATATAATATATAT 286

Qy       1058 TATATTTTCTCTTTGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGT 1117
              ||||| || | | |    |   ||||  |  |   |  |   |  |  | |      ||  |
Db        285 TATATATTTTATATTATATATAATATATATTTTTTATATAATATATATTATATATTTTAT 226

Qy       1118 CATCAACTTTTTCTATATATTATTGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAA 1177
                |  | |  |||||||| ||        | ||  ||  ||    | | || | || |
Db        225 ATATTATATATAATATATATTTTTTATATAATATATATTATATATTTTATATATATTTTA 166
```

```
Qy       1178 CAAAT 1182
              | ||
Db        165 TATAT 161


RESULT 6
US-09-949-002-7549/c
; Sequence 7549, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000790
;  CURRENT APPLICATION NUMBER: US/09/949,002
;  CURRENT FILING DATE:  2000-01-28
;  PRIOR APPLICATION NUMBER: 60/231,401
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 10823
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7549
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-002-7549

  Query Match             3.8%;  Score 46;  DB 3;  Length 601;
  Best Local Similarity   46.8%;  Pred. No. 0.049;
  Matches  178;  Conservative    0;  Mismatches  200;  Indels    2;  Gaps    1;

Qy         47 TATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCC 106
              |||||| | |    |  |  |  ||  |||    |||  |    | |  |  ||||
Db        395 TATTTTATATATTATATATTATTATTATATTATATATTATTTAATATATTATATATTATT 336

Qy        107 CGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTA 166
              || | |||| ||| || |    |       | |||   |   |||||   |  |
Db        335 ATATGTTATTTAATATATTATATATGATATATATKATATTATTTAATATATTATATAT 276

Qy        167 TATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGG 226
              ||    |    |||  ||| ||   |||||    |       |||||| |||||
Db        275 GATATATATTATATATTATTTAATATATTATATATGATATATATTATATATTATTTAATA 216

Qy        227 CAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAA 286
              | | |  ||   ||   | || |   |  ||  |||       ||  || |||||   |
Db        215 TATTATATATGATATATATTATATATTATTTAATATATTATATATTATTTAATATATTAC 156

Qy        287 AAAT--AGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGA 344
              | ||     | |  || ||| ||  | || |   |||| |   |  |||| | | ||  |
Db        155 ATATGATATATATTATATATTTAATATATTACATATGATATATATTATATATTTAATATA 96

Qy        345 TGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGA 404
              |    ||    ||  |   |  ||| |||   |   |||||||    |||||  |    |
Db         95 TTACATATGATATATATTATATATTTAATATATTACATATGATATATATTATATATTTAA 36

Qy        405 ACAATTAAACTTAATTAATA 424
              |||| |  | ||  |||
Db         35 TATATTACATATGATATATA 16


RESULT 7
US-09-949-002-777
; Sequence 777, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000790
;  CURRENT APPLICATION NUMBER: US/09/949,002
;  CURRENT FILING DATE:  2000-01-28
;  PRIOR APPLICATION NUMBER: 60/231,401
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 10823
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 777
;   LENGTH: 25590
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(25590)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-002-777

  Query Match             3.8%;  Score 46;  DB 3;  Length 25590;
  Best Local Similarity   46.8%;  Pred. No. 0.17;
  Matches  178;  Conservative    0;  Mismatches  200;  Indels    2;  Gaps    1;

Qy         47 TATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCC 106
              |||||| | |    |  |  |  ||  |||    |||  |    | |  |  ||||
Db       4195 TATTTTATATATTATATATTATTATTATATTATATATTATTTAATATATTATATATTATT 4254

Qy        107 CGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTA 166
              || | |||| ||| || |    |       | |||   |   |||||   |  |
Db       4255 ATATGTTATTTAATATATTATATATGATATATATTATATATTATTTAATATATTATATAT 4314

Qy        167 TATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGG 226
              ||    |    |||  ||| ||   |||||    |       |||||| |||||
Db       4315 GATATATATTATATATTATTTAATATATTATATATGATATATATTATATATTATTTAATA 4374

Qy        227 CAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAA 286
              | | |  ||   ||   | || |   |  ||  |||       ||  || |||||   |
Db       4375 TATTATATATGATATATATTATATATTATTTAATATATTATATATTATTTAATATATTAC 4434

Qy        287 AAAT--AGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGA 344
              | ||     | |  || ||| ||  | || |   |||| |   |  |||| | | ||  |
Db       4435 ATATGATATATATTATATATTTAATATATTACATATGATATATATTATATATTTAATATA 4494

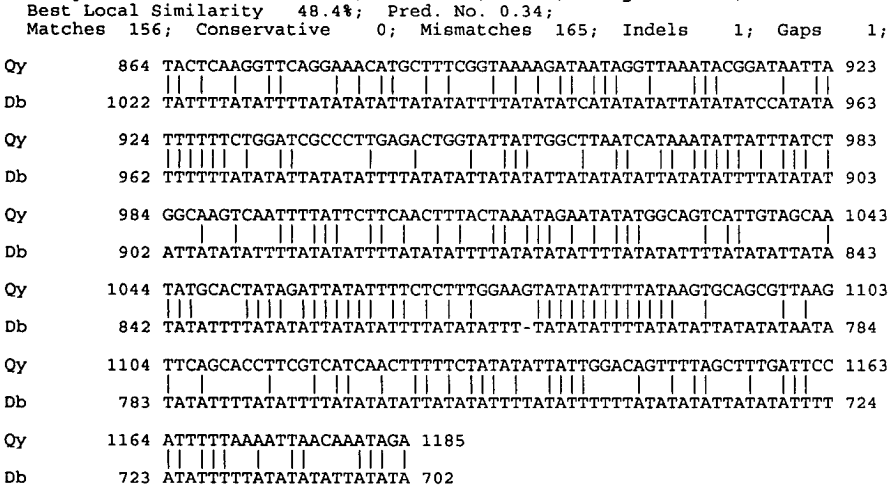
Qy        345 TGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGA 404
              |    ||    ||  |   |  ||| |||   |   |||||||    |||||  |    |
Db       4495 TTACATATGATATATATTATATATTTAATATATTACATATGATATATATTATATATTTAA 4554

Qy        405 ACAATTAAACTTAATTAATA 424
              |||| |  | ||  |||
Db       4555 TATATTACATATGATATATA 4574


RESULT 8
US-09-949-016-12776/c
; Sequence 12776, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12776
;   LENGTH: 187169
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(187169)
;   OTHER INFORMATION: n = A,T,C or G
```

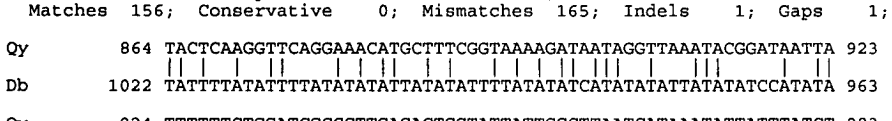
US-09-949-016-12776

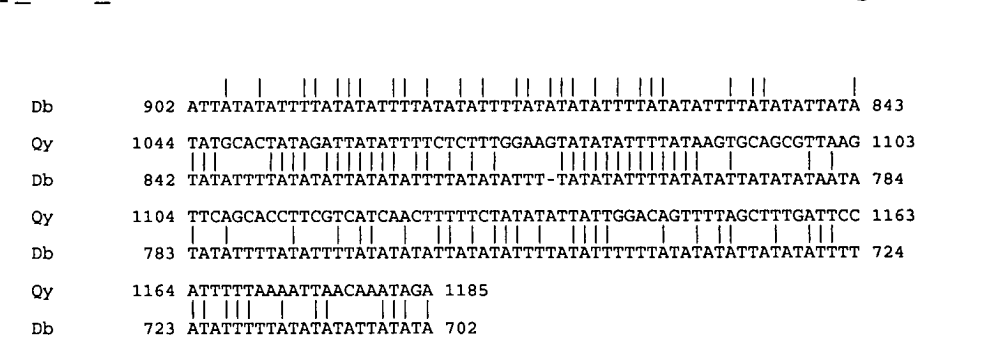
Query Match 3.8%; Score 46; DB 3; Length 187169;
Best Local Similarity 48.4%; Pred. No. 0.34;
Matches 156; Conservative 0; Mismatches 165; Indels 1; Gaps 1;



RESULT 9
US-09-949-016-15940/c
; Sequence 15940, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15940
; LENGTH: 191569
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(191569)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15940

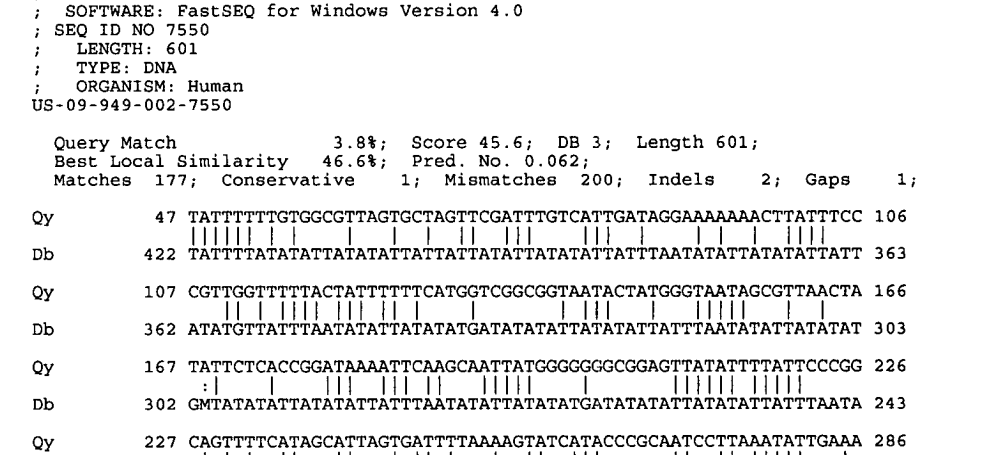
Query Match 3.8%; Score 46; DB 3; Length 191569;
Best Local Similarity 48.4%; Pred. No. 0.34;
Matches 156; Conservative 0; Mismatches 165; Indels 1; Gaps 1;





RESULT 10
US-09-949-002-7550/c
; Sequence 7550, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7550
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-7550

Query Match 3.8%; Score 45.6; DB 3; Length 601;
Best Local Similarity 46.6%; Pred. No. 0.062;
Matches 177; Conservative 1; Mismatches 200; Indels 2; Gaps 1;



RESULT 11
US-09-949-016-17600

```
; Sequence 17600, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17600
;   LENGTH: 50383
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-17600

  Query Match             3.8%;  Score 45.2;  DB 3;  Length 50383;
  Best Local Similarity   48.1%;  Pred. No. 0.34;
  Matches  128;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy        912 TACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGCTTAATCATAAA 971
              ||   |||  ||||| |  || |||     ||  |          ||    ||| ||||
Db      32897 TATTTATATATATTTATAATGTATCTTTGGTGTAAACATGTACATGTATTTAGTCATTTC 32956

Qy        972 TATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATATATGGCAG 1031
              || ||||||| |      |||| |    | ||  | ||     ||||  ||||||  |
Db      32957 TAATATTTATTTAAATTTACAATATGTATTTTATAGTTATTACAATATTATATATGTTAT 33016

Qy       1032 TCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATATATTTTATAAG 1091
              |   | |    ||||  ||||    | ||| |   || | | || | || ||| |||
Db      33017 TGTATATTAATATATTAACTAATATTAATAATGATTCATAGTAATTTTACATTATATTGC 33076

Qy       1092 TGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATTGGACAGTTTT 1151
              |       ||   |     |  |  | |      |||   || || |  ||  | | | |
Db      33077 TTATAAAATATAATTTATATTTAAATAAATCTATTTAAATAAATCTATTTAAATATATAT 33136

Qy       1152 AGCTTTGATTCCATTTTTAAAATTAA 1177
              |  ||     || ||||   | || ||
Db      33137 AAATTATCATCTATTTAATATATAAA 33162


RESULT 12
US-09-949-016-132806/c
; Sequence 132806, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 132806
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-132806

  Query Match             3.8%;  Score 45;  DB 3;  Length 601;
  Best Local Similarity   48.9%;  Pred. No. 0.087;
  Matches  149;  Conservative    0;  Mismatches  155;  Indels    1;  Gaps    1;

Qy        879 AAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGATAATTATTTTTTCTGGATCGC 938
              ||| |   | | |  |||||||| ||  |  | |    |||  |||||| | |   |
Db        559 AAATAGTATCTTGAGAAAAGATACTATATATATTTATTATATATATTTTATATATTTATT 500

Qy        939 CCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCTGGCAAGTCAATTTTA 998
              |   |    ||||||       ||     ||||  ||||| |   |  |  | | ||
Db        499 ATATATAATATATATTATATATAATATATATTATATATTTTATATATTATATATAATATA 440

Qy        999 TTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAG-CAATATGCACTATAGAT 1057
              |  || |  | || || |||  ||||||    |    || ||    ||||  | |||| ||
Db        439 TATTTTATATATAATATATAATATATATTATATATTTTATATATTATATATAATATATAT 380

Qy       1058 TATATTTTCTCTTTGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGT 1117
              ||||| || | | |   |   ||||  |  |   |  |   |  | | |     || |
Db        379 TATATATTTTATATTATATATAATATATATTTTTTATATAATATATATTATATATTTTAT 320

Qy       1118 CATCAACTTTTTCTATATATTATTGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAA 1177
                |  | |  |||||||| ||       | ||  ||  ||    | | || | || |
Db        319 ATATTATATATAATATATATTTTTTATATAATATATATTATATATTTTATATATATTTTA 260

Qy       1178 CAAAT 1182
                | ||
Db        259 TATAT 255


RESULT 13
US-09-949-016-11869
; Sequence 11869, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11869
;   LENGTH: 194714
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-11869

  Query Match             3.8%;  Score 45;  DB 3;  Length 194714;
  Best Local Similarity   48.9%;  Pred. No. 0.61;
  Matches  149;  Conservative    0;  Mismatches  155;  Indels    1;  Gaps    1;

Qy        879 AAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGATAATTATTTTTTCTGGATCGC 938
              ||| |   | | |  |||||||| ||  |  | |    |||  |||||| | |   |
Db      98403 AAATAGTATCTTGAGAAAAGATACTATATATATTTATTATATATATTTTATATATTTATT 98462

Qy        939 CCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCTGGCAAGTCAATTTTA 998
              |   |    ||||||       ||     ||||  ||||| |   |  |  | | ||
Db      98463 ATATATAATATATATTATATATAATATATATTATATATTTTATATATTATATATAATATA 98522

Qy        999 TTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAG-CAATATGCACTATAGAT 1057
              |  || |  | || || |||  ||||||    |    || ||    ||||  | |||| ||
Db      98523 TATTTTATATATAATATATAATATATATTATATATTTTATATATTATATATAATATATAT 98582

Qy       1058 TATATTTTCTCTTTGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGT 1117
              ||||| || | | |   |   ||||  |  |   |  |   |  | | |     || |
Db      98583 TATATATTTTATATTATATATAATATATATTTTTTATATAATATATATTATATATTTTAT 98642
```

```
Qy       1118 CATCAACTTTTTCTATATATTATTGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAA 1177
              |  | |  |||||||| ||       | ||  ||  ||    | | || | || |
Db      98643 ATATTATATATAATATATATTTTTTATATAATATATATTATATATTTTATATATATTTTA 98702

Qy       1178 CAAAT 1182
              | ||
Db      98703 TATAT 98707
```

RESULT 14
US-09-949-016-15474
; Sequence 15474, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15474
; LENGTH: 196714
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15474

Query Match 3.8%; Score 45; DB 3; Length 196714;
Best Local Similarity 48.9%; Pred. No. 0.61;
Matches 149; Conservative 0; Mismatches 155; Indels 1; Gaps 1;

```
Qy        879 AAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGATAATTATTTTTTCTGGATCGC 938
              |||  |   | | |  |||||||| ||  |  | |    |||  |||||| | |   |
Db     148403 AAATAGTATCTTGAGAAAAGATACTATATATATTTATTATATATATTTTATATATTTATT 148462

Qy        939 CCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCTGGCAAGTCAATTTTA 998
              |   |    ||||||       ||     ||||  ||||| |   |  |  | | ||
Db     148463 ATATATAATATATATTATATATAATATATATTATATATTTTATATATTATATATAATATA 148522

Qy        999 TTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAG-CAATATGCACTATAGAT 1057
              |  || |  | || || |||  ||||||   |    || ||    ||||  | |||| ||
Db     148523 TATTTTATATATAATATATAATATATATTATATATTTTATATATTATATATAATATATAT 148582

Qy       1058 TATATTTTCTCTTTGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGT 1117
              ||||| || | | |    |    |||| |  |   |  |  |  |  | |     ||  |
Db     148583 TATATATTTTATATTATATATAATATATATTTTTTATATAATATATATTATATATTTTAT 148642

Qy       1118 CATCAACTTTTTCTATATATTATTGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAA 1177
              |  | |  |||||||| ||       | ||  ||  ||    | | || | || |
Db     148643 ATATTATATATAATATATATTTTTTATATAATATATATTATATATTTTATATATATTTTA 148702

Qy       1178 CAAAT 1182
              | ||
Db     148703 TATAT 148707
```

RESULT 15
US-09-735-271-1038/c
; Sequence 1038, Application US/09735271
; Patent No. 6869762
; GENERAL INFORMATION:
; APPLICANT: Daly, Mark J.
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Lander, Eric S.
; APPLICANT: Rioux, John
; APPLICANT: Siminovitch, Kathy
; TITLE OF INVENTION: IBD-RELATED POLYMORPHISMS
; FILE REFERENCE: 2825.1025-002
; CURRENT APPLICATION NUMBER: US/09/735,271
; CURRENT FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: US 60/170,257
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: US 60/196,046
; PRIOR FILING DATE: 2000-04-10
; NUMBER OF SEQ ID NOS: 2058
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1038
; LENGTH: 700
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(700)
; OTHER INFORMATION: n = A,T,C or G
US-09-735-271-1038

Query Match 3.7%; Score 44.6; DB 3; Length 700;
Best Local Similarity 44.9%; Pred. No. 0.12;
Matches 248; Conservative 0; Mismatches 300; Indels 4; Gaps 2;

```
Qy        144 ACTATGGGTAATAGCGTTAACTATATTCTCACCGGATAAAATTCAAGCAATTATGGGGGG 203
              | |||   | |     |||| ||||| | |    || ||  |  |   | |||
Db        684 AATATATATTAATATATTAAATATATATTAATATAATTAATATATATTTAATATAATNTA 625

Qy        204 GCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATA 263
              | | |  | ||||          |       ||||   |||| |||  |||  ||
Db        624 ATATAATAAATATATATTATATATTAAATAATATATATTATATATTTAAAATATATATTA 565

Qy        264 CCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGT 323
              ||  |  |   || ||| |     |   ||   ||  | | |||  | |||  |
Db        564 ATATATATTATATATATTTAAAATATATAATATATATTAAACATATTTAAAATATATATT 505

Qy        324 TGCGACAATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATA 383
              |   | |||||  |||||  ||    ||    ||  |   |  |||| ||       |||
Db        504 T--AAAAATATAAATTATATATTATATATAATATATTATATATTATATATAATATATA 447

Qy        384 TGATGAGCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCG 443
              ||    | |||  |  | |  |||| |  ||||  |||     ||   ||  |
Db        446 ATATATATAATATATAATATATATTATATATAATATATTATATATTATATATTTAT 387

Qy        444 GGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTT 503
              |  |     | |   |    | |   | |  |||||  |      |   | | ||
Db        386 ATATTTATATATTGTATATTTATATATTGTATATATTTATATATTATATATTTATATATT 327

Qy        504 GTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTT 563
              |||| |||    | | |||    |   |||| |   |  || ||   |||| |||  |
Db        326 TTTATATATTATATATATTTATATATATTATATATATT--TTTATATATATTTTTATATA 269

Qy        564 TGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCT 623
              |   ||||||||  || |  ||   | ||  || |     ||| || |  |  |
Db        268 TTATATATTTATATATATTTTTATATATTTTTATATATTTTATATATTTATATATTTTT 209

Qy        624 TATTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATT 683
              ||| ||     | ||| ||  |  |  |     | ||  | | |  |   |||  ||||
Db        208 TATATATTATATATATTTATATATATTTTTATATATTATATATATTTATATATATTTATA 149

Qy        684 TGTAATAATTAT 695
              | | |||  |||
Db        148 TATTATATATAT 137
```

Search completed: May 12, 2006, 03:51:52
Job time : 169.123 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:31:18 ; Search time 768.343 Seconds
(without alignments)
12915.142 Million cell updates/sec

Title: US-09-423-093B-3_COPY_5114_6313
Perfect score: 1200
Sequence: 1 atgctgccatttccaccagg.....atagacggtgcacgcgataa 1200

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1200	100.0	12441	3	US-09-423-093A-3	Sequence 3, Appli
2	58.6	4.9	18154	6	US-10-311-455-227	Sequence 227, App
3	53.4	4.5	16750	6	US-10-311-455-494	Sequence 494, App
4	53.4	4.5	16750	7	US-10-221-714A-36	Sequence 36, Appl
5	51.4	4.3	16373	6	US-10-311-455-591	Sequence 591, App
6	50.6	4.2	6161	7	US-10-221-613-160	Sequence 160, App
7	50.6	4.2	16766	6	US-10-311-455-2130	Sequence 2130, Ap
8	50	4.2	5536	6	US-10-311-455-152	Sequence 152, App
9	49.6	4.1	6381	6	US-10-311-455-679	Sequence 679, App
10	49.6	4.1	3673778	6	US-10-312-841-2	Sequence 2, Appli
11	49.2	4.1	3089	9	US-10-954-778-160	Sequence 160, App
12	49.2	4.1	4985	5	US-10-094-240-10	Sequence 10, Appl
13	49.2	4.1	4985	6	US-10-056-405-10	Sequence 10, Appl
14	49.2	4.1	6220	6	US-10-311-455-1273	Sequence 1273, Ap
15	49.2	4.1	7319	6	US-10-311-455-2018	Sequence 2018, Ap
16	49.2	4.1	8222	9	US-10-486-319A-43	Sequence 43, Appl
17	49.2	4.1	8222	9	US-10-486-319A-65	Sequence 65, Appl
18	49.2	4.1	8771	6	US-10-311-455-1797	Sequence 1797, Ap
19	49.2	4.1	9504	6	US-10-240-453-281	Sequence 281, App
20	49	4.1	3673778	6	US-10-312-841-1	Sequence 1, Appli
21	48.8	4.1	7317	5	US-10-239-676-47	Sequence 47, Appl
22	48.8	4.1	7317	6	US-10-240-453-47	Sequence 47, Appl
23	48.8	4.1	8260	6	US-10-240-453-215	Sequence 215, App
24	48.8	4.1	13202	6	US-10-311-455-1458	Sequence 1458, Ap
25	48.6	4.0	8845	7	US-10-221-714A-266	Sequence 266, App
26	48.6	4.0	11394	6	US-10-240-453-95	Sequence 95, Appl
27	48.4	4.0	3051	9	US-10-486-319A-67	Sequence 67, Appl
28	48.4	4.0	8962	6	US-10-311-455-659	Sequence 659, App
29	48	4.0	5511	6	US-10-311-455-1844	Sequence 1844, Ap
30	48	4.0	6754	7	US-10-221-613-266	Sequence 266, App
31	48	4.0	8413	6	US-10-240-485-50	Sequence 50, Appl
32	48	4.0	9731	6	US-10-311-455-964	Sequence 964, App
33	48	4.0	17491	6	US-10-240-485-128	Sequence 128, App
c 34	47.8	4.0	408	9	US-10-275-323A-13	Sequence 13, Appl
35	47.8	4.0	6239	7	US-10-221-613-248	Sequence 248, App
36	47.6	4.0	8530	6	US-10-311-455-1728	Sequence 1728, Ap
37	47.6	4.0	9095	7	US-10-433-793-92	Sequence 92, Appl
38	47.4	4.0	48265	7	US-10-741-601-5677	Sequence 5677, Ap
39	47.2	3.9	5888	6	US-10-240-485-10	Sequence 10, Appl
40	47.2	3.9	6063	6	US-10-240-453-267	Sequence 267, App
41	47.2	3.9	6392	6	US-10-311-455-657	Sequence 657, App
42	47.2	3.9	6392	6	US-10-240-485-59	Sequence 59, Appl
43	47.2	3.9	17594	6	US-10-311-455-1999	Sequence 1999, Ap
44	47	3.9	5407	6	US-10-311-455-2064	Sequence 2064, Ap
45	47	3.9	6378	6	US-10-311-455-150	Sequence 150, App

ALIGNMENTS

RESULT 1
US-09-423-093A-3
; Sequence 3, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12441
; TYPE: DNA
; ORGANISM: Salmonella enterica
US-09-423-093A-3

Query Match 100.0%; Score 1200; DB 3; Length 12441;
Best Local Similarity 100.0%; Pred. No. 1.1e-250;
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGCTGCCATTTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5114 ATGCTGCCATTTCCACCAGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCG 5173

Qy         61 TTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5174 TTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACT 5233

Qy        121 ATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5234 ATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGAT 5293

Qy        181 AAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5294 AAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCA 5353

Qy        241 TTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5354 TTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTAC 5413
```

```
Qy        301 ATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5414 ATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGA 5473

Qy        361 GAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5474 GAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATT 5533

Qy        421 AATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5534 AATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGA 5593

Qy        481 TATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5594 TATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGA 5653

Qy        541 TTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5654 TTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGA 5713

Qy        601 GGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5714 GGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATG 5773

Qy        661 CTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5774 CTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAAT 5833

Qy        721 ATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5834 ATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCG 5893

Qy        781 CGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAGAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5894 CGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAGAA 5953

Qy        841 CATCCATCAGGTATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5954 CATCCATCAGGTATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGAT 6013

Qy        901 AATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6014 AATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGC 6073

Qy        961 TTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6074 TTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGA 6133

Qy       1021 ATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6134 ATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATA 6193

Qy       1081 TATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6194 TATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTAT 6253

Qy       1141 TGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGACGGTGCACGCGATAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6254 TGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGACGGTGCACGCGATAA 6313
```

RESULT 2
US-10-311-455-227
; Sequence 227, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 227
; LENGTH: 18154
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-227

Query Match 4.9%; Score 58.6; DB 6; Length 18154;
Best Local Similarity 48.4%; Pred. No. 0.063;
Matches 193; Conservative 0; Mismatches 204; Indels 2; Gaps 1;

```
Qy        350 TAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAAT 409
              |||   | | |  | | || |  ||  | | |||| |  | |||    | | |  | |
Db      13388 TAAAATAGGAATGGATAATGATTTTTATTGAATATTTTTTGTATTTATGAAAATGATTAA 13447

Qy        410 TAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTC 469
              | |  ||   |  ||  ||      |   |  ||| |  |   |||||| |  | |
Db      13448 TTATTTTTTATTTTAAATAATTATTGTATTTTTGGGATTAATATTGTTTTATTTTTTTTA 13507

Qy        470 TCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAG 529
              | |||| | ||  |||| | |   ||  |  ||| |||| |    | |  ||||   |||
Db      13508 TAAATTGTTGAATTATGTTTAATATATTTTATTTATTATTTTAAAATTTGTTTTTAGAAG 13567

Qy        530 GATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGA 589
              |  || ||| ||||  |  || || ||  |||  || ||   ||    | |     |
Db      13568 TAATTATAAGTTTATTTTTATTTTTTTTTGTTTTGATGTTAAAATTATATTTGTTTTATA 13627

Qy        590 GTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAA 649
              | | | | | ||   |  || ||  |   ||||| |    |||||||| | ||
Db      13628 AAATGATTTGGGTAGTTTTTTTTATTTTTTTTTTTATTAAAATTATTTATATAAGATTAG 13687

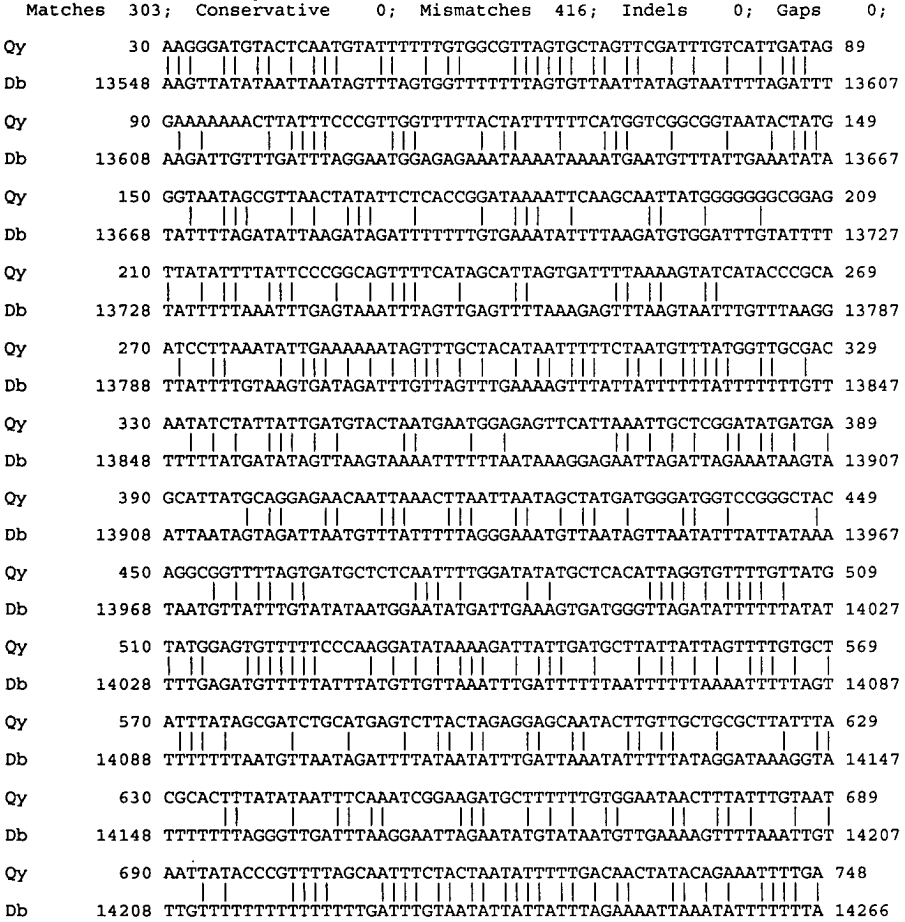
Qy        650 ATCGGAAGATGCTTT--TTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGC 707
              |     ||  |||    | | |   || ||| |||||||| | |    || |.|
Db      13688 AATTATTTATTGTTTGAAGTTTTGGTAAAATTTGTTTGTAATATTATTTAGTTTTGTTGT 13747

Qy        708 AATTTCTACTAATATTTTTGACAACTATACAGAAATTTT 746
              |||    |   |  |||  ||  |  |  ||| ||
Db      13748 TTTTTGAGGTGGGAGATTTTTGAATATTGAATTAATATT 13786
```

RESULT 3
US-10-311-455-494
; Sequence 494, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 494
; LENGTH: 16750
; TYPE: DNA

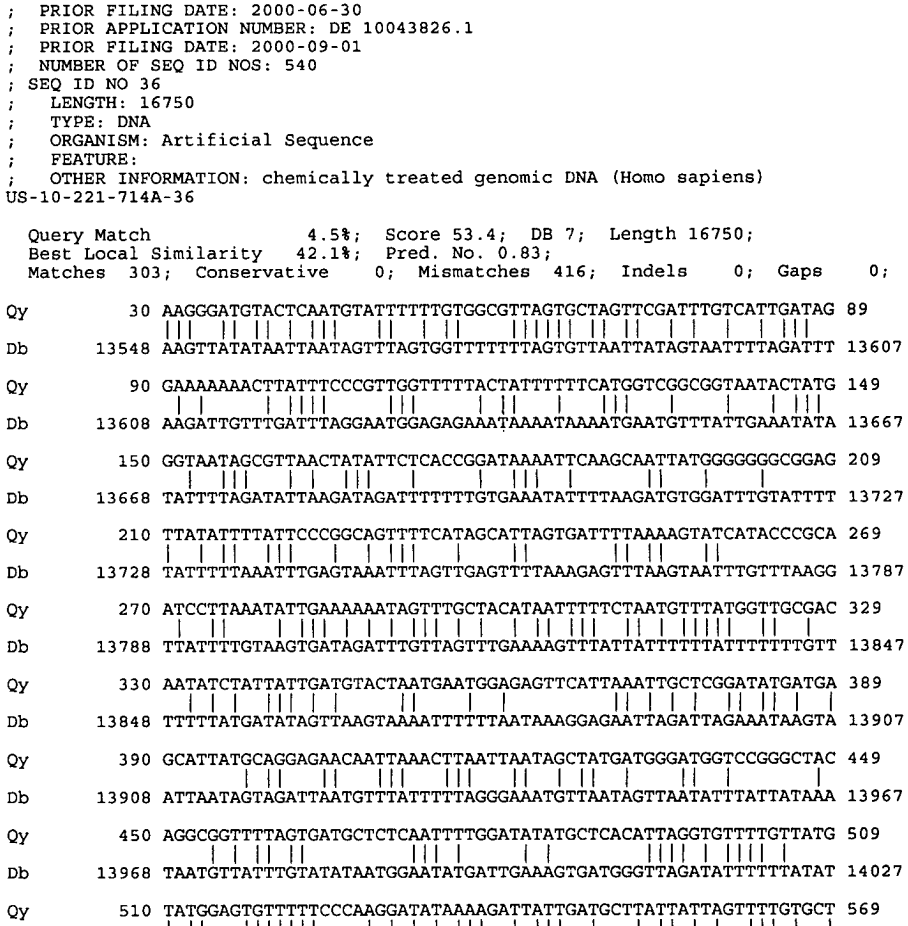
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-494

Query Match 4.5%; Score 53.4; DB 6; Length 16750;
Best Local Similarity 42.1%; Pred. No. 0.83;
Matches 303; Conservative 0; Mismatches 416; Indels 0; Gaps 0;



RESULT 4
US-10-221-714A-36
; Sequence 36, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 36
; LENGTH: 16750
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-36

Query Match 4.5%; Score 53.4; DB 7; Length 16750;
Best Local Similarity 42.1%; Pred. No. 0.83;
Matches 303; Conservative 0; Mismatches 416; Indels 0; Gaps 0;



RESULT 5

```
US-10-311-455-591
; Sequence 591, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 591
;   LENGTH: 16373
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-591

  Query Match             4.3%;  Score 51.4;  DB 6;  Length 16373;
  Best Local Similarity   44.8%;  Pred. No. 2.2;
  Matches  246;  Conservative    0;  Mismatches  296;  Indels    7;  Gaps    1;

Qy        206 GGAGTTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACC 265
              ||| |||  | |      ||  ||  | |||  |||  | |||||  |    ||  ||
Db       8709 GGATTTAAGTGTATAAAATGGAGGTGATAATAATATTTTTTATTTTGAGTTGATGTTAGA 8768

Qy        266 CGCAATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTG 325
              | |  |||   |||||  | | || |||| ||  |||  ||| | || | ||    |
Db       8769 TTTAGTTGTTATTAATTGATAGAGTATTTTGATAATTAAAATTTTTGATTTATAAATATT 8828

Qy        326 CGACAATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATG 385
              |    |  |||| || ||       | ||   ||  |   || |||| |   |
Db       8829 TTATTTAAGATATTTTTAATTATAAGTTTAAAATAGTTTAATTTTAATTGGTTTAAATGA 8888

Qy        386 ATGAGCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGG 445
              |  |  | ||| |   |       ||     ||||        |  |  | | |  |||
Db       8889 AAAAATAGAATGTATTGGTTGTTATATTAGGAATTTTATAAAGTTTTTTGGTAGGGTGGG 8948

Qy        446 CTACAGGCGGTTTTA-------GTGATGCTCTCAATTTTGGATATATGCTCACATTAGGT 498
              | |||     ||      ||  |  | | | ||||  |||  |  |    ||  ||
Db       8949 TTGTAGGTATAGTTTTATTAGGGTATTAGTTTTATTTTTTTATAGTTTTTTTAGTTTTGT 9008

Qy        499 GTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATT 558
              |||| |||| ||| || | | |||        |||    |  |||| | |  |   | ||
Db       9009 TTTTTTTTATATATTGATTTTATTTGTAGGTAATAGGTTAAGTTATAGGTTTTATATGTT 9068

Qy        559 AGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCT 618
              | || |  |||||||||  |     |  | | ||  |  || |     || | |  |
Db       9069 TATATTATATTATTTATAGTAAGAGAGAAAATTTTTTTTTTAGAATTTTTAGTAGAAGTA 9128

Qy        619 GCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACT 678
              |      ||   |  |||     |     |  |  ||||   |||    |
Db       9129 TTAAGATGTGTTTTGATTGGATTGGTTTGTGGTATATGAGTTTTAGTTTGAAAAATTGAT 9188

Qy        679 TTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACA 738
              |||  |     | ||       |||| | |  || |      |||||  | |  ||||
Db       9189 TTAATATAGGTTATATGTTTTTTTTTTGGATATTGTTTAGCGATTTTATAGATTTATAGT 9248

Qy        739 GAAATTTTG 747
              |  |||| |
Db       9249 GGTATTTAG 9257


RESULT 6
```

```
US-10-221-613-160
; Sequence 160, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, ALexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
;  FILE REFERENCE: 5013.1004
;  CURRENT APPLICATION NUMBER: US/10/221,613
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: PCT/EP01/02945
;    DE 10013847.00
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-03-15
;    2000-03-15
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 160
;   LENGTH: 6161
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-160

  Query Match             4.2%;  Score 50.6;  DB 7;  Length 6161;
  Best Local Similarity   43.6%;  Pred. No. 2.2;
  Matches  318;  Conservative    0;  Mismatches  409;  Indels    2;  Gaps    2;

Qy         18 AGGAGCAATCCTAAGGGATGTACTCAATGTATTTTTTGTGGCGTTAGTGCTAGTTCGATT 77
              | ||  ||   |||   |  |  |     || ||  | || | || |||    ||     |
Db       2326 AAGATAAAGTTTAATTTAGTTGTTTTTGGTTTTAGTAGTTGGGTAAGTAAAGGTAAAGGT 2385

Qy         78 TGTCATTGATAGGAAAAAAACTTATTTCCCGTTGGTTTTTACTATTTTTTCATGGTCGGC 137
                 ||  |   ||  | || || |||     || |||||  | |     |  |||   |
Db       2386 AAAGATAAAGGCGAGTATAATTTTTTTTTTTTTTGTTTTAATTTAAATCGGATGTCGGTT 2445

Qy        138 GGTAATACTATGGGTAATAGCGTTAACTATATTCTCACCGGATAAAATTCAAGCAATTAT 197
              |   |  | |   |    | | ||  |   ||   |     | |   |  |  || | |
Db       2446 AGATTTTTTTTTTTTGGGTGTGGTAGGTGGTTTACGATGTTTTGAGTGTGTATAAAGTGT 2505

Qy        198 GGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGT 257
              |||| |  | ||| | |||||||  ||  | ||||         ||  |    |   |||
Db       2506 TGGGGAGGAGGGTTTTGTTTTATTTTCGTTAATTTTTGGGAATTTTTTTTAGTATTTAGT 2565

Qy        258 ATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTT 317
              |  ||          |   |  |  |||||       |   || |   |  ||
Db       2566 TTTGTAGTTTTTGAGGAAATTATTATAGTAAATATAAGATTGTGTATTGAATTAAAGTGA 2625

Qy        318 TATGGTTGCGACAATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCT 377
              || |  |      ||  | | | ||   |   | |||  | |  | |     | |  | |
Db       2626 TAGGAATTTAGAGATTTTTTTGATAAGTTGTTTTATGGTTAGTAAATAGGGGATAGGGAT 2685

Qy        378 CGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGAT 437
                | | ||     ||||  |         |||      |  ||     |  || | |  |
Db       2686 -TTTTTTTATTTTTTTTATAGAAAGATGTAATATTGTATTTTTTTAGGTAATAAAGTATT 2744

Qy        438 GGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGG 497
              |||         |     |||  ||    | |   ||||    |   |  |     | |||
Db       2745 GGTATTTTGGGAATTAAAATTTTTTGGGTGTTTTTTTTTTTTTTTCGTTTTATATTAAGG 2804

Qy        498 TGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTAT 557
              | ||| | | | | |    | |||||          | | | |  |||||   |  |||||
Db       2805 TATTTAGGTTTATTTTTTTTTTTTTTTTTATTTTTTTTTATATTTTATTTTTTTTTATTTT 2864

Qy        558 TAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGC 617
```

```
Db       2865 TAGTTTTATGGTTTTTTATTAGTAGGGAGTTGGGGAATAGTTAAGTTTTTTAAGTTTTTTT 2924

Qy        618 TGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAAC 677

Db       2925 TAGGGAAGTTTAGGCGATTTTGAGAAATATAAGTTGAAA-ATGTAATGGAAGGGAATAGT 2983

Qy        678 TTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATAC 737

Db       2984 TTTTTATAGAAATGTTATAGTTAAGAGATATATTTTTAATATGGATTTAAAGAATATTTT 3043

Qy        738 AGAAATTTT 746

Db       3044 AGAAATTAT 3052


RESULT 7
US-10-311-455-2130
; Sequence 2130, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2130
;   LENGTH: 16766
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2130

  Query Match             4.2%;  Score 50.6;  DB 6;  Length 16766;
  Best Local Similarity   45.2%;  Pred. No. 3.4;
  Matches  447;  Conservative    0;  Mismatches  514;  Indels   28;  Gaps    6;

Qy        208 AGTTATATTTTATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCG 267

Db       7054 ATTTATGTGGTTTTTTAGGTTGAATTTTAATTTTGTTATAGGTTTATTTTGTGATATTGT 7113

Qy        268 CAATCCTTAAATATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCG 327

Db       7114 GGTTTTTTTTATTTTGAAAGAAGAGTTTTATTTGTTATTTATTTTGAGATTAGATAAATT 7173

Qy        328 ACAATATCTATTATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGAT 387

Db       7174 TTGTTAATTTTTAAGGTTTAGATAAAATGGATGGATTGAGTGAAGTGATTTTTTGAATTT 7233

Qy        388 GAGCATTATGCAGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCT 447

Db       7234 TAATTTTTTTTAGATAGATAAATAATTTTTTAGGAATTCGTTGTTTTATAGTTTTTTGGT 7293

Qy        448 ACAGGCGGTTTTAGTGATGCTCTCAATTTTGGATATATGCTCACAT---------TAGGT 498

Db       7294 AAAGTATGGTTGTGTTTTTTTTTTTGTTTGATATAGATTGATGTATATTGTTAAGGAGGT 7353

Qy        499 GTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATT 558

Db       7354 AATGAATAATGAATGAAGTGGATTTAGAAATGTTAGATAAAA------ATAATTATATAT 7407

Qy        559 AGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCT 618

Db       7408 TTTTTTTTGGAAATTAGAAAAATATGAAAAAGTATAAATAAGAAAGTAATATTTATTTTT 7467

Qy        619 GCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACT 678

Db       7468 ATTTTTA-TTATTTAGTATTAATAATTGTTAATATTTTGGTATATTTGATTTTAGTTTAT 7526

Qy        679 TTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACA-ACTATAC 737

Db       7527 TTTTTAAATATTTTTTTTTTTTTTTGTAAAAATTATTATTATTATTAATATTAGGCGA 7586

Qy        738 AGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCATCGCGTGGATCTACACAGGG 797

Db       7587 ATATAATTTTATAGATTTTTTTATGTATAAATATATATATATATATATATATATATATAT 7646

Qy        798 GCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAGAACATCCATCAGGTATAGG 857

Db       7647 ATATATATATAGTGTTTTTGTTTAATGTTATTTGGGAAAAGATTAAAATAAAGTTAAAGG 7706

Qy        858 TCTGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGA 917

Db       7707 AGTTGTTAA--------AATTTAAATTTTTTTGTTTATAAGAAAGTATAATTTTTTATGA 7758

Qy        918 TAATTATTTT---TTCTGGATCGCCCTTGAGACTGGTATTATTGGCTTAATCATAAATAT 974

Db       7759 TATTTTTTTAAGGTTTTAAATAAAATTTATGAAAGATATTAAAAGTTGTAGTATTATATA 7818

Qy        975 TATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATATATGGCAGTCA 1034

Db       7819 TTTATATAGTATTTTTTAATTTTATATTATATTTATTATTTTATTTTTTGAAAGAAGAGA 7878

Qy       1035 TTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATATATTTTTATAAGTGC 1094

Db       7879 ATATTGGTATTTTTATGTTTTTTTTTATTTTTTTTTTTGATTTTTTTTTTTTTTGGTTGT 7938

Qy       1095 AGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATTGGACAGTTTTAGC 1154

Db       7939 ATATTTTTGTTTAGATTAATAATAATCGTATTTTATTTTATTTATAATTTTTATAGATGT 7998

Qy       1155 TTTGATTCCATTTTTAAAATTAACAAATA 1183

Db       7999 ATTATTTTTATATTTAAATAGAATAAATA 8027


RESULT 8
US-10-311-455-152
; Sequence 152, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 152
;   LENGTH: 5536
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-152

  Query Match             4.2%;  Score 50;  DB 6;  Length 5536;
```

```
  Best Local Similarity   46.3%;  Pred. No. 2.8;
  Matches  164;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;

Qy        357 TGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTT 416
              | |||  || |||  |||  | | | |  | |||   ||       | |  |  ||| |
Db       2919 TTGAGTTTTAATTTTATTATTTGTAAAATAGGAGTTATAATGGTAATTATGAGGAAATGT 2978

Qy        417 AATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTT 476
              || | |    | |     | |  |  |    | |    ||||| |   ||  | | ||
Db       2979 AAGTGACGTTTGTAGAATGGTTTTTGGTATAATAAAGAGTTTTTGATTTGTGTTATAGTT 3038

Qy        477 TGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAA 536
              || || ||   |  ||| |  ||||   |  | | || ||| |||    |   | |
Db       3039 TGTATGTAAAATTTTATTTGTATTTTTTAGAATTGTTGAATGTGTTTTGAGAAATTTTGG 3098

Qy        537 AAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTAC 596
              |         ||||| | |   |    | | | |      | ||  | | | |||
Db       3099 TTTTATTGGTGGATTTATTAATTGGAGTTGTGTTTATTTTATTTTTTGTTTTATTTTTAA 3158

Qy        597 TAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAA 656
              |  |      | || || ||  |     |||       | ||| |||| |||  ||| |
Db       3159 TTTAATGAAATTATTTTTTTTTAGAGATATGATTTGATTTTTTATATTTTAGAATTGATG 3218

Qy        657 GATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAAT 710
              |   | | | ||  ||| | |||| | | ||||||||     ||||  | | |
Db       3219 TAGTTTATATGGTTAAATTATTTTAATAGAAATAATTAAGAATGTTTATGAATT 3272


RESULT 9
US-10-311-455-679
; Sequence 679, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 679
;   LENGTH: 6381
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-679

  Query Match             4.1%;  Score 49.6;  DB 6;  Length 6381;
  Best Local Similarity   51.3%;  Pred. No. 3.7;
  Matches  115;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy        481 TATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGA 540
              ||| || | |  || ||| |     | |||||     |||||| |    | || ||  ||
Db       3091 TATTTGATAATTTTGGGTATAGGAATTTGTATTTTAGGTTTTTACGGTGAATTTAGTAGT 3150

Qy        541 TTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTAGA 600
              | |||  |      || ||   |||||  |||||||| || |  | ||  |  ||| |
Db       3151 TAATTTTTAGAGTTTGTTTTGTTTGTTTTATTTATATCGTTTGGTATATGGGTTAGTTCG 3210

Qy        601 GGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATG 660
                |  || ||  |       ||| || |  ||||||  |||     |  |
Db       3211 TTTTTATTTATTTTTTTTATTTGGTTTTTCGGATGTTATATTTTTTGGTTTTTGGTTGGT 3270

Qy        661 CTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTT 704


              ||||| || |  |   || |||| || |  |  ||    ||||
Db       3271 TTTTTTAGTTGGTTTTTTTAATTTATATTTGTAGTATGGATTTT 3314


RESULT 10
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (379615)
US-10-312-841-2

  Query Match             4.1%;  Score 49.6;  DB 6;  Length 3673778;
  Best Local Similarity   55.2%;  Pred. No. 58;
  Matches  139;  Conservative    0;  Mismatches  109;  Indels    4;  Gaps    2;

Qy        499 GTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATT 558
              || ||  ||| | ||  || |||||   ||| || |||| |   || |    || |||||
Db    1083532 GTGTTTATATTTGTGATGTTTTTTTTAGGAAGAATTTAAATGTAAATAGTATTTTTTTATT 1083591

Qy        559 AGTTTTGTGCT-ATTTATAGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGC 617
              |||||| |  |  ||| |    || |  ||  |  ||| ||||     ||      ||
Db    1083592 AGTTTTATTTTGTTTTTTTTATATTTTAATTGGAATTATTAGATTTATAAGTGAGTTTTT 1083651

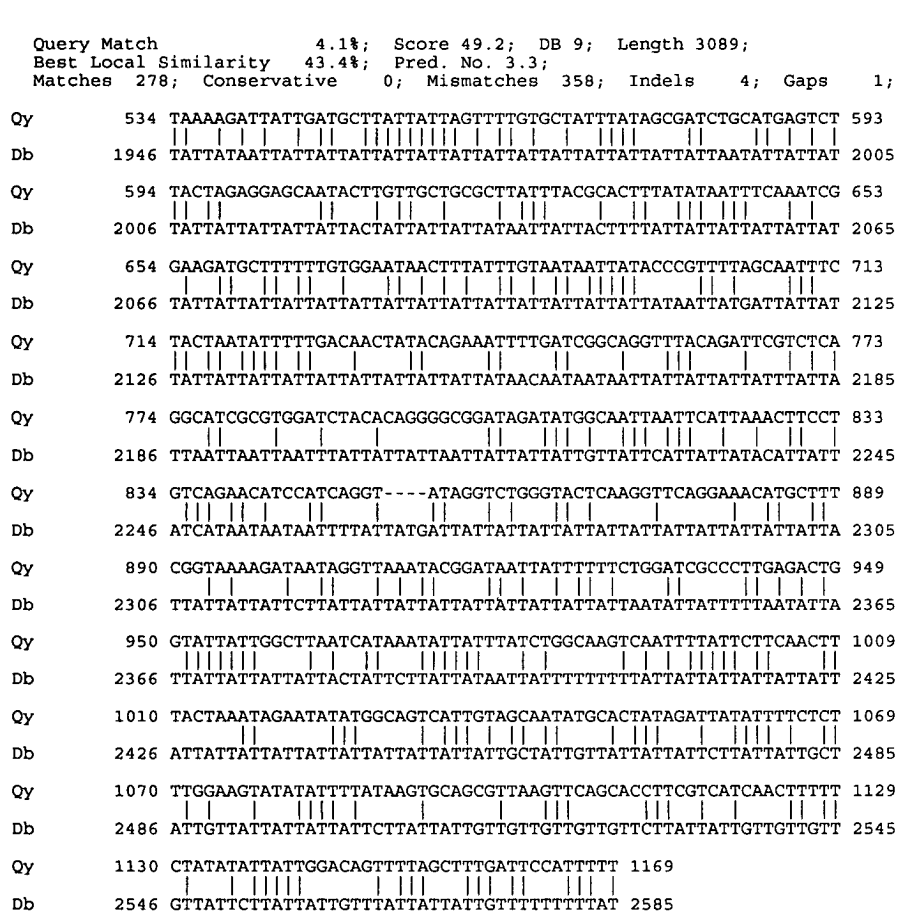
Qy        618 TGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAAC 677
              |     | |||    |  || | |  |||    | |||    |  |||||  |   ||
Db    1083652 TTTTTGTTTTTTTTAATGTTTTTTGGTTTTTTGTAGGATATAGGGTTTTTTTTATAT--- 1083708

Qy        678 TTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATAC 737
              |||| |||| || ||| |      || ||| | ||| || | |||||||| |    |||
Db    1083709 TTTAATTGTTATTATTTTTATTTTTGTAGGATTTTTTATTTATATTTTTTAATTGTATTT 1083768

Qy        738 AGAAATTTTGAT 749
              | |  |||||||
Db    1083769 AAATTTTTTGAT 1083780

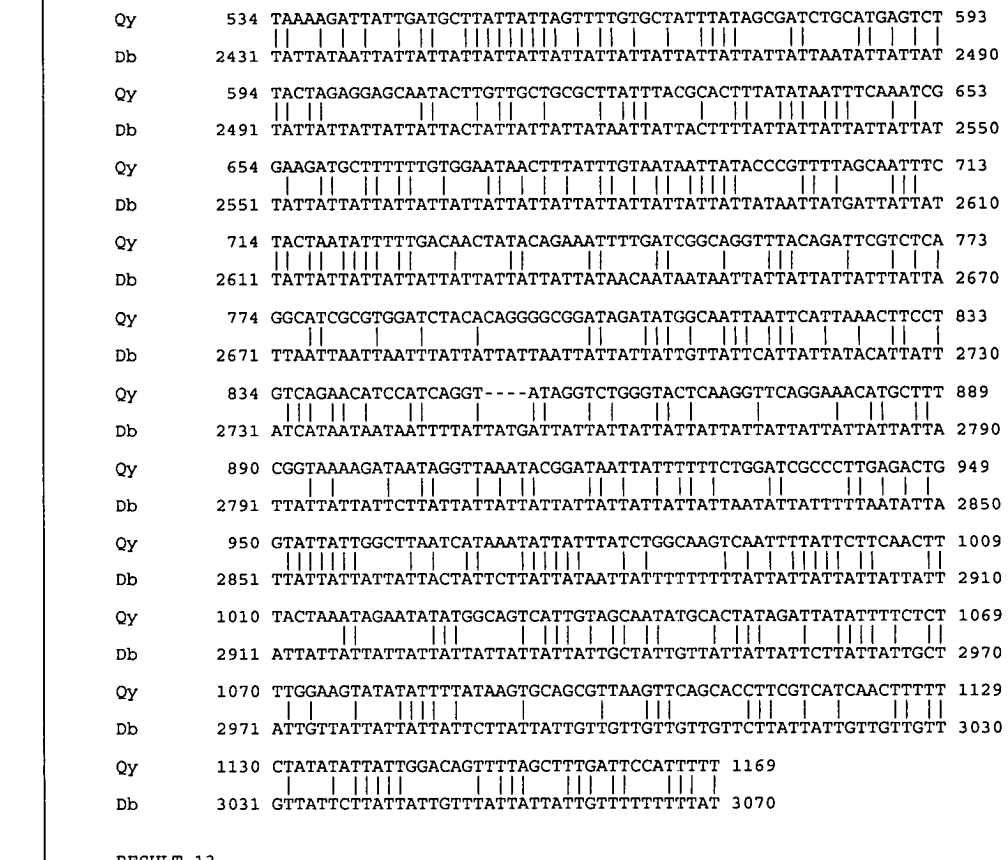

RESULT 11
US-10-954-778-160
; Sequence 160, Application US/10954778
; Publication No. US20050153368A1
; GENERAL INFORMATION:
;  APPLICANT: ZWIEBEL, LAURENCE J.
;  TITLE OF INVENTION: METHOD OF IDENTIFYING CHEMICAL AGENTS WHICH STIMULATE
;  TITLE OF INVENTION:  ODORANT RECEPTORS OF SENSORY NEURONS
;  FILE REFERENCE: N1125
;  CURRENT APPLICATION NUMBER: US/10/954,778
;  CURRENT FILING DATE:  2004-09-30
;  PRIOR APPLICATION NUMBER: 10/056,405
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: 60/264,649
;  PRIOR FILING DATE: 2001-01-26
;  NUMBER OF SEQ ID NOS: 248
;  SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 160
;   LENGTH: 3089
;   TYPE: DNA
;   ORGANISM: Anopheles gambiae
US-10-954-778-160
```

Query Match 4.1%; Score 49.2; DB 9; Length 3089;
Best Local Similarity 43.4%; Pred. No. 3.3;
Matches 278; Conservative 0; Mismatches 358; Indels 4; Gaps 1;



RESULT 12
US-10-094-240-10
; Sequence 10, Application US/10094240
; Publication No. US20030082637A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
; FILE REFERENCE: N8289
; CURRENT APPLICATION NUMBER: US/10/094,240
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 10/056,405
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-094-240-10

Query Match 4.1%; Score 49.2; DB 5; Length 4985;
Best Local Similarity 43.4%; Pred. No. 4;
Matches 278; Conservative 0; Mismatches 358; Indels 4; Gaps 1;



RESULT 13
US-10-056-405-10
; Sequence 10, Application US/10056405
; Publication No. US20030166013A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: MOSQUITO OLFACTORY GENES, POLYPEPTIDES, AND METHODS OF
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: N7841
; CURRENT APPLICATION NUMBER: US/10/056,405
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-056-405-10

Query Match 4.1%; Score 49.2; DB 6; Length 4985;
Best Local Similarity 43.4%; Pred. No. 4;
Matches 278; Conservative 0; Mismatches 358; Indels 4; Gaps 1;

```
Qy        534 TAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCT 593
              ||  | | |  | ||  |||||||||| | || |  |  ||||    ||    || | | |
Db       2431 TATTATAATTATTATTATTATTATTATTATTATTATTATTATTATTAATATTATTAT 2490

Qy        594 TACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCG 653
              ||  ||          ||   | ||  |   | |||     |  ||  ||| |||    | ||
Db       2491 TATTATTATTATTATTACTATTATTATTATAATTATTACTTTTATTATTATTATTATTAT 2550

Qy        654 GAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTC 713
              |  ||  || ||  |   || | | |  || | ||  |||||     || |     |||
Db       2551 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATAATTATGATTATTAT 2610

Qy        714 TACTAATATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCA 773
              || ||  |||| ||    |    ||      ||     |   |||     |   | | |
Db       2611 TATTATTATTATTATTATTATTATTATTATAACAATAATAATTATTATTATTATTTATTA 2670

Qy        774 GGCATCGCGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCT 833
              ||    |   |       |          ||   ||| |  ||| |||  |   |  ||  |
Db       2671 TTAATTAATTAATTTATTATTATTAATTATTATTATTGTTATTCATTATTATACATTATT 2730

Qy        834 GTCAGAACATCCATCAGGT----ATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCTTT 889
              |||  ||  |   ||    |   ||   | |   || |      |      |  ||  ||
Db       2731 ATCATAATAATAATTTTATTATGATTATTATTATTATTATTATTATTATTATTATTA 2790

Qy        890 CGGTAAAAGATAATAGGTTAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTG 949
              | |     |  ||   | | ||    || | |  | ||  |    ||     || | | |
Db       2791 TTATTATTATTCTTATTATTATTATTATTATTATTATTATTAATATTATTTTTAATATTA 2850

Qy        950 GTATTATTGGCTTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTT 1009
              |||||||    | |  ||   ||||||   | |      | | | ||||| ||    ||
Db       2851 TTATTATTATTATTACTATTCTTATTATAATTATTTTTTTTTATTATTATTATTATTATT 2910

Qy       1010 TACTAAATAGAATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCT 1069
                ||      |||     | ||| | || ||    | |||   |  |||| |  ||
Db       2911 ATTATTATTATTATTATTATTATTATTATTGCTATTGTTATTATTATTCTTATTATTGCT 2970

Qy       1070 TTGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTT 1129
              | |   |  ||||  |      |      |   |||      |||  |  |    || | |
Db       2971 ATTGTTATTATTATTATTCTTATTATTGTTGTTGTTGTTGTTCTTATTATTGTTGTTGTT 3030

Qy       1130 CTATATATTATTGGACAGTTTTAGCTTTGATTCCATTTTT 1169
              |   | |||||      | |||   ||| ||   ||| |
Db       3031 GTTATTCTTATTATTGTTTATTATTATTGTTTTTTTTTAT 3070
```

RESULT 14
US-10-311-455-1273
; Sequence 1273, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1273
; LENGTH: 6220
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 920, 1012
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1273

Query Match 4.1%; Score 49.2; DB 6; Length 6220;
Best Local Similarity 48.9%; Pred. No. 4.4;
Matches 132; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

```
Qy        456 TTTTAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGA 515
              |||||||| || | | |   || | | |||       ||||||  |   |  |||||  |
Db       3502 TTTTAGTGGTGATTTGAGAGTTTGGTGTATTTATTATTTAGGTAGTGGATATTGTATTTA 3561

Qy        516 GTGTTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTAT 575
              ||| |   |       | |  |    | || ||  |  || |||   || |   |||| |
Db       3562 ATGTGTAGTCGTTTATTTTTTATTTATTTTTATTTTAGTTTTTATAATTTTAAGATTTTT 3621

Qy        576 AGCGATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACT 635
              |  || |       | ||| |           |  ||  |  |    || ||||   | |
Db       3622 GGTTATTTAGGGTTATTTTATTTTTTTTTTTTTTTTTTGTTGTATTTTTTTTTATTTATT 3681

Qy        636 TTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTAT 695
              |||| |  |||     |   |  | | || ||| ||| |  |   |  || |||| |||
Db       3682 TTATTTTTTTATTTTTTAAGTAAGATTATTTTTGGTAAGATGGTTTTGGTAAGAATGTG 3741

Qy        696 ACCCGTTTTAGCAATTTCTACTAATATTTT 725
              |  ||| |||  |   | |  |||  ||||
Db       3742 ATACGTATTATAATAATTTTTTAAATTTTT 3771
```

RESULT 15
US-10-311-455-2018
; Sequence 2018, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2018
; LENGTH: 7319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2018

Query Match 4.1%; Score 49.2; DB 6; Length 7319;
Best Local Similarity 43.9%; Pred. No. 4.7;
Matches 302; Conservative 0; Mismatches 383; Indels 3; Gaps 2;

```
Qy         39 ACTCAATGTATTTTTTGTGGCGTTAGTGCTAGTTCGATTTGTCATTGATAGGAAAAAAAC 98
              | | ||  | || | | |    ||  || |   |  |||| | |||  |  | |
Db       5866 ATTTAAATTGTTATATTTTTTTTTTTTGATGAATTTATTTTTTATTATTTAGTAGTTGTT 5925

Qy         99 TTATTTCCCGTTGGTTTTTACTATTTTTTCATGGTCGGCGGTAATACTATGGGTAATAGC 158
              || ||    ||  | ||   | || ||| |   |   | |||   |   |||||||
Db       5926 TTTTTGATTATTAATATTGTTTTTTGTTTTAAAATTTTTGTTAAATATTAAGGTAATATT 5985

Qy        159 GTTAACTATATTCTCACCGGATAAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTT 218
              |    | | || |    |  |  | || |   | || |  |          || | |||
Db       5986 TTAGTTTTTTTTTTATTGGTGTTTATTTTATATATTTTTTCGTTTTTTATTTTTTTTTTA 6045
```

```
Qy        219 ATTCCCGGCAGTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAA 278
              |||        |||||  |    ||      || ||||  |  | ||   | | |   |
Db       6046 ATTTTTTTTTGTTTTTTTTATTTTGTGTTTTATAAATTTTACGTAGTTGGATTTTGTTTT 6105

Qy        279 TATTGAAAAAATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTAT 338
              | |     || || ||    ||  |  |  ||   | ||||||  |    | | |   | |
Db       6106 TTTATTTAATATGGTAGTATAATTTTTAATTGGTAAGTTTATTTTATTTATATTTATTGT 6165

Qy        339 TATTGATGTACTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGC 398
              |||  ||   ||    |    |||       | |  |   |  |  |      || |
Db       6166 AATTATTGATATATATTTGGATTTGTTTGAATTAATTTTATTAGTTTTTTTTTTTTTTTA 6225

Qy        399 AGGAGAACAATTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTT 458
              |        ||      |  |||| |   |||  |||   ||  |   ||       |||
Db       6226 TGTTTTTTTTTTTTTGTTTTTAAAAAAAATGTCGGGTATGTGTGTTTTATTTTTTTTTT 6285

Qy        459 TAGTGATGCTCTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTG 518
              |  |     | |   || |  || | |  |     |||  | |    || ||| | |  |
Db       6286 TTTTTTAAATTTATGTTATTAATTTTTTTTAGTGTTATTTTTAAATTTTTGTTTTGTATA 6345

Qy        519 TTTTTCCCAAGGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGC 578
              |||     ||  |  |   |  |||   ||  | | ||||| ||||    |||||  |
Db       6346 TTTAATATAATAAAGTTTTAAGTTAAATATTTTAAATATTATTTTTAGATAATTTAAAAA 6405

Qy        579 GATCTGCATGAGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTA 638
              |  | ||  || |||   | |||     |  || | | |     |||  | |   | |
Db       6406 TTTTAGAAT--GTTTTAAGTGTGGATATTTTTTTTTAGATTTTTATATAGATG-TTTATG 6462

Qy        639 TATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATACC 698
              |||   |  | ||       ||   | | | || | |   |||||||  | ||  || |
Db       6463 TATGTGTGTATATATATGTATTTATATGTTTGTATTTTATTTATTTATTATTTTTTTTTA 6522

Qy        699 CGTTTTAGCAATTTCTACTAATATTTTT 726
                ||||||    || ||  | |||||||
Db       6523 TATTTTAGATTATTTTATGATTATTTTT 6550
```

Search completed: May 12, 2006, 05:34:58
Job time : 791.343 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:52:36 ; Search time 611.399 Seconds
(without alignments)
8004.956 Million cell updates/sec

Title: US-09-423-093B-3_COPY_5114_6313
Perfect score: 1200
Sequence: 1 atgctgccatttccaccagg.....atagacggtgcacgcgataa 1200

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/2/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/2/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	53	4.4	843	12	US-10-301-480-545203	Sequence 545203,
	2	53	4.4	843	12	US-10-301-480-1158612	Sequence 1158612,
	3	53	4.4	851	12	US-10-301-480-545202	Sequence 545202,
	4	53	4.4	851	12	US-10-301-480-1158611	Sequence 1158611,
	5	53	4.4	875	12	US-10-301-480-545201	Sequence 545201,
	6	53	4.4	875	12	US-10-301-480-1158610	Sequence 1158610,
	7	53	4.4	938	12	US-10-301-480-545200	Sequence 545200,
	8	53	4.4	938	12	US-10-301-480-1158609	Sequence 1158609,
c	9	48.6	4.0	687411	11	US-10-330-773-26	Sequence 26, Appl
c	10	47.6	4.0	119036	10	US-10-995-561-13314	Sequence 13314, A
c	11	47	3.9	50959	17	US-11-117-187-210	Sequence 210, App
c	12	47	3.9	72600	17	US-11-117-187-206	Sequence 206, App
	13	46.6	3.9	9347	10	US-10-240-708-35	Sequence 35, Appl
	14	46.6	3.9	171486	17	US-11-121-086-105	Sequence 105, App
	15	46	3.8	852	12	US-10-301-480-567933	Sequence 567933,
	16	46	3.8	852	12	US-10-301-480-1181342	Sequence 1181342,
	17	45.2	3.8	638	12	US-10-301-480-574942	Sequence 574942,
	18	45.2	3.8	638	12	US-10-301-480-1188351	Sequence 1188351,
	19	44.8	3.7	1313	17	US-11-136-527-745	Sequence 745, App
	20	44.8	3.7	1313	17	US-11-136-527-4841	Sequence 4841, Ap
	21	44.8	3.7	20674	17	US-11-124-367A-5035	Sequence 5035, Ap
c	22	44.6	3.7	541	7	US-09-925-065A-616669	Sequence 616669,
c	23	44.6	3.7	1281	10	US-10-750-185-47170	Sequence 47170, A
c	24	44.6	3.7	1281	10	US-10-750-623-47170	Sequence 47170, A
	25	44.4	3.7	145985	11	US-10-330-773-140	Sequence 140, App
	26	44.2	3.7	12872	10	US-10-995-561-13511	Sequence 13511, A
	27	43.8	3.6	579	7	US-09-925-065A-566811	Sequence 566811,
c	28	43.2	3.6	319608	17	US-11-145-703-1	Sequence 1, Appli
c	29	43	3.6	590	11	US-10-301-480-56052	Sequence 56052, A
c	30	43	3.6	590	12	US-10-301-480-669461	Sequence 669461,
	31	43	3.6	6040	10	US-10-240-708-70	Sequence 70, Appl
	32	42.8	3.6	642	7	US-09-925-065A-785322	Sequence 785322,
c	33	42.8	3.6	643	7	US-09-925-065A-786236	Sequence 786236,
	34	42.8	3.6	647	7	US-09-925-065A-739823	Sequence 739823,
	35	42.8	3.6	173602	17	US-11-121-086-25	Sequence 25, Appl
	36	42.6	3.6	579	11	US-10-301-480-82990	Sequence 82990, A
	37	42.6	3.6	579	11	US-10-301-480-82991	Sequence 82991, A
	38	42.6	3.6	579	12	US-10-301-480-696399	Sequence 696399,
	39	42.6	3.6	579	12	US-10-301-480-696400	Sequence 696400,
	40	42.6	3.6	19233	10	US-10-240-708-46	Sequence 46, Appl
	41	42.6	3.6	1457619	17	US-11-098-686-8739	Sequence 8739, Ap
c	42	42.4	3.5	201	10	US-10-995-561-53061	Sequence 53061, A
c	43	42.4	3.5	613	12	US-10-301-480-363169	Sequence 363169,
c	44	42.4	3.5	613	12	US-10-301-480-976578	Sequence 976578,
c	45	42.4	3.5	1128	7	US-09-925-065A-286218	Sequence 286218,

ALIGNMENTS

RESULT 1
US-10-301-480-545203
; Sequence 545203, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 545203
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-545203

Query Match 4.4%; Score 53; DB 12; Length 843;
Best Local Similarity 53.1%; Pred. No. 1.7;
Matches 113; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```
Qy       964 ATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATA 1023
             ||  || |||| || |||||     |     || ||||   ||   | || || |||  |||
Db       476 ATATTATATATAATATATCTATATATAATATCTTATATATCTTATATATTATATATTATA 535

Qy      1024 TATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATATAT 1083
             |||      |  |        |||| |  |||    |||||||| | | | ||  |||||||
Db       536 TATATATATTTTATATATTATATACTATATTTTATATATTTTATATATTGATATATATAT 595

Qy      1084 TTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATTGG 1143
              |||||  |      |||  |  |  |  |   | |  |  | |||  ||||| ||||
```

```
Db        596 ATTATATATTATATATTATATATAATATATAAATAAATATATATTTATATATTTTATATA 655

Qy       1144 ACAGTTTTAGCTTTGATTCCATTTTTAAAATTA 1176
              | |  |  |  | |   |  || ||  | ||||
Db        656 ATATATAAAAATATATATTTATATTATATATTA 688


RESULT 2
US-10-301-480-1158612
; Sequence 1158612, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1158612
;   LENGTH: 843
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1158612

  Query Match             4.4%;  Score 53;  DB 12;  Length 843;
  Best Local Similarity   53.1%;  Pred. No. 1.7;
  Matches  113;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        964 ATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATA 1023
              ||  || |||| || |||||     |    || ||||   ||   | || || |||  |||
Db        476 ATATTATATATAATATATCTATATATAATATCTTATATATCTTATATATTATATATTATA 535

Qy       1024 TATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATATAT 1083
              |||      |  |       |||| |  |||    |||||||| | | | ||  |||||||
Db        536 TATATATATTTTATATATTATATACTATATTTTATATATTTTATATATTGATATATATAT 595

Qy       1084 TTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATTGG 1143
              |||||  |       |||  |  |  |   | |  |  | ||| |||||  ||||
Db        596 ATTATATATTATATATTATATATAATATATAAATAAATATATATTTATATATTTTATATA 655

Qy       1144 ACAGTTTTAGCTTTGATTCCATTTTTAAAATTA 1176
              | |  |  |  | |   |  || ||  | ||||
Db        656 ATATATAAAAATATATATTTATATTATATATTA 688


RESULT 3
US-10-301-480-545202
; Sequence 545202, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 545202
;   LENGTH: 851
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-545202

  Query Match             4.4%;  Score 53;  DB 12;  Length 851;
  Best Local Similarity   53.1%;  Pred. No. 1.7;
  Matches  113;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        964 ATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATA 1023
              ||  || |||| || |||||     |    || ||||   ||   | || || |||  |||
Db        484 ATATTATATATAATATATCTATATATAATATCTTATATATCTTATATATTATATATTATA 543

Qy       1024 TATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATATAT 1083
              |||      |  |       |||| |  |||    |||||||| | | | ||  |||||||
Db        544 TATATATATTTTATATATTATATACTATATTTTATATATTTTATATATTGATATATATAT 603

Qy       1084 TTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATTGG 1143
              |||||  |       |||  |  |  |   | |  |  | ||| |||||  ||||
Db        604 ATTATATATTATATATTATATATAATATATAAATAAATATATATTTATATATTTTATATA 663

Qy       1144 ACAGTTTTAGCTTTGATTCCATTTTTAAAATTA 1176
              | |  |  |  | |   |  || ||  | ||||
Db        664 ATATATAAAAATATATATTTATATTATATATTA 696


RESULT 4
US-10-301-480-1158611
; Sequence 1158611, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1158611
;   LENGTH: 851
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1158611

  Query Match             4.4%;  Score 53;  DB 12;  Length 851;
  Best Local Similarity   53.1%;  Pred. No. 1.7;
  Matches  113;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        964 ATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATA 1023
              ||  || |||| || |||||     |    || ||||   ||   | || || |||  |||
Db        484 ATATTATATATAATATATCTATATATAATATCTTATATATCTTATATATTATATATTATA 543

Qy       1024 TATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATATAT 1083
              |||      |  |       |||| |  |||    |||||||| | | | ||  |||||||
Db        544 TATATATATTTTATATATTATATACTATATTTTATATATTTTATATATTGATATATATAT 603

Qy       1084 TTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATTGG 1143
              |||||  |       |||  |  |  |   | |  |  | ||| |||||  ||||
Db        604 ATTATATATTATATATTATATATAATATATAAATAAATATATATTTATATATTTTATATA 663

Qy       1144 ACAGTTTTAGCTTTGATTCCATTTTTAAAATTA 1176
              | |  |  |  | |   |  || ||  | ||||
Db        664 ATATATAAAAATATATATTTATATTATATATTA 696


RESULT 5
US-10-301-480-545201
; Sequence 545201, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
```

```
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 545201
;   LENGTH: 875
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-545201

  Query Match             4.4%;  Score 53;  DB 12;  Length 875;
  Best Local Similarity   53.1%;  Pred. No. 1.7;
  Matches  113;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        964 ATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATA 1023
              ||  || |||| || |||||    |    || ||||   ||   | || || |||  |||
Db        508 ATATTATATATAATATATCTATATATAATATCTTATATATCTTATATATTATATATTATA 567

Qy       1024 TATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATATAT 1083
              |||      |  |        |||| |  |||     |||||||| | | | ||  |||||||
Db        568 TATATATATTTTATATATTATATACTATATTTTATATATTTTATATATTGATATATATAT 627

Qy       1084 TTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATTGG 1143
               |||||  |       |||  |  |  |  |    | |  |  | ||| ||||| ||||
Db        628 ATTATATATTATATATTATATATAATATATAAATAAATATATATTTATATATTTTATATA 687

Qy       1144 ACAGTTTTAGCTTTGATTCCATTTTTAAAATTA 1176
              | |  |  |  | |   |  || ||  | ||||
Db        688 ATATATAAAAATATATATTTATATTATATATTA 720


RESULT 6
US-10-301-480-1158610
; Sequence 1158610, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1158610
;   LENGTH: 875
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-1158610

  Query Match             4.4%;  Score 53;  DB 12;  Length 875;
  Best Local Similarity   53.1%;  Pred. No. 1.7;
  Matches  113;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        964 ATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATA 1023
              ||  || |||| || |||||    |    || ||||   ||   | || || |||  |||
Db        508 ATATTATATATAATATATCTATATATAATATCTTATATATCTTATATATTATATATTATA 567

Qy       1024 TATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATATAT 1083
              |||      |  |        |||| |  |||     |||||||| | | | ||  |||||||
Db        568 TATATATATTTTATATATTATATACTATATTTTATATATTTTATATATTGATATATATAT 627

Qy       1084 TTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATTGG 1143
               |||||  |       |||  |  |  |  |    | |  |  | ||| ||||| ||||
Db        628 ATTATATATTATATATTATATATAATATATAAATAAATATATATTTATATATTTTATATA 687

Qy       1144 ACAGTTTTAGCTTTGATTCCATTTTTAAAATTA 1176
              | |  |  |  | |   |  || ||  | ||||
Db        688 ATATATAAAAATATATATTTATATTATATATTA 720


RESULT 7
US-10-301-480-545200
; Sequence 545200, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 545200
;   LENGTH: 938
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-545200

  Query Match             4.4%;  Score 53;  DB 12;  Length 938;
  Best Local Similarity   53.1%;  Pred. No. 1.7;
  Matches  113;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        964 ATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATA 1023
              ||  || |||| || |||||    |    || ||||   ||   | || || |||  |||
Db        571 ATATTATATATAATATATCTATATATAATATCTTATATATCTTATATATTATATATTATA 630

Qy       1024 TATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATATAT 1083
              |||      |  |        |||| |  |||     |||||||| | | | ||  |||||||
Db        631 TATATATATTTTATATATTATATACTATATTTTATATATTTTATATATTGATATATATAT 690

Qy       1084 TTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATTGG 1143
               |||||  |       |||  |  |  |  |    | |  |  | ||| ||||| ||||
Db        691 ATTATATATTATATATTATATATAATATATAAATAAATATATATTTATATATTTTATATA 750

Qy       1144 ACAGTTTTAGCTTTGATTCCATTTTTAAAATTA 1176
              | |  |  |  | |   |  || ||  | ||||
Db        751 ATATATAAAAATATATATTTATATTATATATTA 783


RESULT 8
US-10-301-480-1158609
; Sequence 1158609, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1158609
;   LENGTH: 938
;   TYPE: DNA
;   ORGANISM: Homo sapien
```

US-10-301-480-1158609

Query Match 4.4%; Score 53; DB 12; Length 938;
Best Local Similarity 53.1%; Pred. No. 1.7;
Matches 113; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```
Qy        964 ATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATA 1023
              ||  || |||| || |||||     |    || ||||   ||   | || || |||  |||
Db        571 ATATTATATATAATATATCTATATATAATATCTTATATATCTTATATATTATATATTATA 630

Qy       1024 TATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATATAT 1083
              |||     |  |        |||| |  |||    |||||||| | | | ||  |||||||
Db        631 TATATATATTTTATATATTATATACTATATTTTATATATTTTATATATTGATATATATAT 690

Qy       1084 TTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATTGG 1143
              |||||  |     |||  |  |  |  |  | |  |  | ||| |||||  ||||
Db        691 ATTATATATTATATATTATATATAATATATAAATAAATATATATTTATATATTTTATATA 750

Qy       1144 ACAGTTTTAGCTTTGATTCCATTTTTAAAATTA 1176
              | |  |  |  | |   |  || ||  | ||||
Db        751 ATATATAAAAATATATATTTATATTATATATTA 783
```

RESULT 9
US-10-330-773-26/c
; Sequence 26, Application US/10330773
; Publication No. US20060040262A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001300
; CURRENT APPLICATION NUMBER: US/10/330,773
; CURRENT FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 981
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 687411
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(687411)
; OTHER INFORMATION: n = A,T,C or G
US-10-330-773-26

Query Match 4.0%; Score 48.6; DB 11; Length 687411;
Best Local Similarity 43.0%; Pred. No. 31;
Matches 237; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

```
Qy        635 TTTATATAATTTCAAATCGGAAGATGCTTTTTTGTGGAATAACTTTATTTGTAATAATTA 694
              |||||||| || ||||    | ||  ||   | | | | |    ||  || |  ||
Db        635 TTTATATATATTTAAATATATATATATTTAAATATATATTTATATATATTTAAAAATATA 576

Qy        695 TACCCGTTTTAGCAATTTCTACTAATATTTTTGACAACTATACAGAAATTTTGATCGGCA 754
              ||    ||| |  |  |  |  ||    ||| | |  |||   | || |   |     |
Db        575 TATATATTTAAATATATATTTATATATATTTAAATATATATATTTATATATATTTAAAAA 516

Qy        755 GGTTTACAGATTCGTCTCAGGCATCGCGTGGATCTACACAGGGGCGGATAGATATGGCAA 814
              | || | |||              |  || ||    |      ||| || |
Db        515 TATATATATATTTAAAAATATATATATTTATATATATTTATATAAATATATATTTATATT 456

Qy        815 TTAATTCATTAAACTTCCTGTCAGAACATCCATCAGGTATAGGTCTGGGTACTCAAGGTT 874
              |  ||    || |  |  |  | ||           ||||    |    | |  |  ||
Db        455 TATATAATATATATATATTTATAAAAATATATATTAATATATAAATATTAAGTATATATT 396

Qy        875 CAGGAAACATGCTTTCGGTAAAAGATAATAGGTTAAATACGGATAATTATTTTTTCTGGA 934
              |   ||      | |   ||   ||  |||  |  || |   | |||   |||||    |
Db        395 TATATAAATATATGTATATATTTGAATATATATGTAAGATATAAAATACATTTTTAAATA 336

Qy        935 TCGCCCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCTGGCAAGTCAAT 994
              |    ||     |  |  || ||   | ||  || ||||||| ||| |  | |  | ||
Db        335 TTTTTATTATATTTTATTATAATGAAATTATATTATATATTATATATTTTACTATACTAT 276
```

```
Qy        995 TTTATTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAGCAATATGCACTATA 1054
              ||| | |  ||  |     | || |||||   | | |      |   ||||  | ||||
Db        275 ATTAATATAAAATATATTGTATTACAATATGCTGTATTATAATTTAATATATATAATATA 216

Qy       1055 GATTATATTTTCTCTTTGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGCACCTT 1114
              | |||||    |  |    | | ||  |   |||      |    |||  |  |  |  |
Db        215 TAATATATAAAATTATATATAATTTAATTACAATATAATTATATATAATTATAATATAAT 156

Qy       1115 CGTCATCAACTTTTTCTATATATTATTGGACAGTTTTAGCTTTGATTCCATTTTTAAAAT 1174
              |||    | ||| |||      ||     |||    |    |     || || |||
Db        155 ATACATTTTATATTTATATTATAAATAAATATTTTTACTATAATTTAATTATTATATAAT 96

Qy       1175 TAACAAATAGA 1185
              ||  | | | |
Db          95 TATAATAAATA 85
```

RESULT 10
US-10-995-561-13314/c
; Sequence 13314, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13314
; LENGTH: 119036
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(119036)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-995-561-13314

Query Match 4.0%; Score 47.6; DB 10; Length 119036;
Best Local Similarity 50.4%; Pred. No. 36;
Matches 113; Conservative 1; Mismatches 110; Indels 0; Gaps 0;

```
Qy        962 TAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAA 1021
              ||||| ||  |  | | ||| |   |  | |  | || || |  |||  || ||| |
Db      31817 TAATCTTATTTTATTTATATTTCAAATATATAAAATCTTATTTACATTTCATATATATAT 31758

Qy       1022 TATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTTTGGAAGTATAT 1081
              | ||   | | |:  |    ||||  |  | |  |||| || | | |||  |  |||||
Db      31757 AAAATCTTATTTAYATTTCATATATATATAAAATCTTATTTTATATATTTTTATATATAT 31698

Qy       1082 ATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTCTATATATTATT 1141
              || ||||  |  |   |||  ||    |  ||     ||  | ||| | ||||||  ||
Db      31697 TTTATATATATAAAATCTTATTTTATATATATTTTATATATATTTTATATATATAAAATC 31638

Qy       1142 GGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGA 1185
                |   | |     |  | ||   ||||| |  ||    ||| |
Db      31637 ATATTTTATATATATAAAATCTTATTTTATATATATTTTATATA 31594
```

RESULT 11
US-11-117-187-210/c
; Sequence 210, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE
; APPLICANT: COPENHAVER, GREGORY
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD:309US
; CURRENT APPLICATION NUMBER: US/11/117,187

```
;  CURRENT FILING DATE:  2005-04-28
;  PRIOR APPLICATION NUMBER: US/09/531,120
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: 60/125,219
;  PRIOR FILING DATE: 1999-03-18
;  NUMBER OF SEQ ID NOS: 212
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 210
;   LENGTH: 50959
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-11-117-187-210

  Query Match             3.9%;  Score 47;  DB 17;  Length 50959;
  Best Local Similarity   52.9%;  Pred. No. 40;
  Matches  101;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy        951 TATTATTGGCTTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTT 1010
              | || ||   |||||| || | | | |  | |   | ||| || |||     |   |  ||
Db      17263 TTTTTTTTTTTTTAATTATGAGTTTGAAGTTTGAAGAAAGACATTTTCCAAGTATTAAATT 17204

Qy       1011 ACTAAATAGAATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTT 1070
              |  ||||   |||  | ||   |||  ||   | || ||| | |   || |  | || |
Db      17203 ATCAAATTTTATAATTAGCTTACATGTTAATGACATACACGAAACTGTACAAATACTATG 17144

Qy       1071 TGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTC 1130
              || |    || || |  ||| |  |   ||| ||||   | | |    | ||||| |
Db      17143 TGTACACCTACATATACTAAATATACAATTACGTTCCAAAACGTACATACCAACTGTAGA 17084

Qy       1131 TATATATTATT 1141
                |||||| ||
Db      17083 CATATATTTTT 17073


RESULT 12
US-11-117-187-206/c
; Sequence 206, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
;  APPLICANT: PREUSS, DAPHNE
;  APPLICANT:  COPENHAVER, GREGORY
;  TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
;  FILE REFERENCE: ARCD:309US
;  CURRENT APPLICATION NUMBER: US/11/117,187
;  CURRENT FILING DATE:  2005-04-28
;  PRIOR APPLICATION NUMBER: US/09/531,120
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: 60/125,219
;  PRIOR FILING DATE: 1999-03-18
;  NUMBER OF SEQ ID NOS: 212
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 206
;   LENGTH: 72600
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-11-117-187-206

  Query Match             3.9%;  Score 47;  DB 17;  Length 72600;
  Best Local Similarity   52.9%;  Pred. No. 43;
  Matches  101;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy        951 TATTATTGGCTTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTT 1010
              | || ||   |||||| || | | | |  | |   | ||| || |||     |   |  ||
Db      17263 TTTTTTTTTTTTTAATTATGAGTTTGAAGTTTGAAGAAAGACATTTTCCAAGTATTAAATT 17204

Qy       1011 ACTAAATAGAATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTT 1070
              |  ||||   |||  | ||   |||  ||   | || ||| | |   || |  | || |
Db      17203 ATCAAATTTTATAATTAGCTTACATGTTAATGACATACACGAAACTGTACAAATACTATG 17144

Qy       1071 TGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTC 1130
              || |    || || |  ||| |  |   ||| ||||   | | |    | ||||| |
Db      17143 TGTACACCTACATATACTAAATATACAATTACGTTCCAAAACGTACATACCAACTGTAGA 17084

Qy       1131 TATATATTATT 1141
                |||||| ||
Db      17083 CATATATTTTT 17073


RESULT 13
US-10-240-708-35
; Sequence 35, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/240,708
;  CURRENT FILING DATE:  2002-10-03
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 35
;   LENGTH: 9347
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-35

  Query Match             3.9%;  Score 46.6;  DB 10;  Length 9347;
  Best Local Similarity   44.4%;  Pred. No. 36;
  Matches  187;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy        229 GTTTTCATAGCATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAA 288
              ||||   ||| |||| |  ||||   ||||  | |     ||  | |     |   ||
Db       1553 GTTTATTTAGTATTAATTTTTTTTTTAGTAATAAAATTTGTATATTGATGGGGTAGTAAT 1612

Qy        289 ATAGTTTGCTACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTA 348
              |  |||  ||   |  || |   ||  || |       | ||      || || |  |
Db       1613 GTGTTTTTTTAATAATATTATAATATAATTTTTAAATTTATAAGGAGGTTTTTTTAGTTT 1672

Qy        349 CTAATGAATGGAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAA 408
              |  | || | |  |||   |  |       | | |      ||  | ||||     |
Db       1673 TTTTTAAAGGTATGGTTATGTGGTTAAAGTATAAGTAAAAGTGGTTGGGAAGGATTTTTA 1732

Qy        409 TTAAACTTAATTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCT 468
              ||  ||  ||  ||   |||    |||       || | |  ||||  || | |
Db       1733 GAAAGATTTTTTTTTAAAAATGGATAGATAACGGTTTGTAAAAGTTATTTTTTTGTTTTT 1792

Qy        469 CTCAATTTTGGATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAA 528
              |   ||||  ||| | |    ||   | |||| || | |         ||||   ||
Db       1793 TTTTTTTTTTTTTATTTATTTTTTTTTTATTTTTTTTTTTTGTTTTATGTTTTAAAA 1852

Qy        529 GGATATAAAAGATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATG 588
              |||      || ||  |  ||||| || ||| | ||  ||  |  | ||  |  |||
Db       1853 TTTTATTTGTTTTTTTTTTTTTTATTTTTCGTTATTTGGAATGGAATGTGAATTTGATG 1912

Qy        589 AGTCTTACTAGAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCA 648
              |  | |  || ||  ||  | | |  || | | |||       |  | || | |||
Db       1913 GTTGGAATTTTAGAAGTTATTTTGGATAATGAGGTGATTGTAAGGATGGAGATTAATTCG 1972

Qy        649 A 649
              |
Db       1973 A 1973
```

```
RESULT 14
US-11-121-086-105
; Sequence 105, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 105
;   LENGTH: 171486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-105

  Query Match             3.9%;  Score 46.6;  DB 17;  Length 171486;
  Best Local Similarity   42.5%;  Pred. No. 58;
  Matches  431;  Conservative    0;  Mismatches  574;  Indels   10;  Gaps    3;

Qy        179 ATAAAATTCAAGCAATTATGGGGGGGCGGAGTTATATTTTATTCCCGGCAGTTTTCATAG 238
              |||| ||  ||   ||||||          ||||||||||        | |
Db      16549 ATAATATATAATATATTATGTATTATATATTATATATTTTATATAATATTATATATTATT 16608

Qy        239 CATTAGTGATTTTAAAAGTATCATACCCGCAATCCTTAAATATTGAAAAAATAGTTTGCT 298
              || |   || | || |||  ||      | |  ||| ||| | | || |||  |
Db      16609 ATATAATATATTAATAAATATATTATATATATTTATTATATAGT-ATAATATATTACATA 16667

Qy        299 ACATAATTTTTCTAATGTTTATGGTTGCGACAATATCTATTATTGATGTACTAATGAATG 358
              |  ||  |   || || ||  |    |||||  |    | || || ||   |||
Db      16668 TTAATATAATAAATATATTAATTATATAATAAATATATTAATATAATATATTATATAATA 16727

Qy        359 GAGAGTTCATTAAATTGCTCGGATATGATGAGCATTATGCAGGAGAACAATTAAACTTAA 418
              | |  | | | |||       ||   ||  |  |||   |    |   || | || | |
Db      16728 AATATATTAATTAATATAATATATTATATATGTCTTAAATATATAACATATAATACATTA 16787

Qy        419 TTAATAGCTATGATGGGATGGTCCGGGCTACAGGCGGTTTTAGTGATGCTCTCAATTTTG 478
              ||       |          || ||     |      |  |       ||| |
Db      16788 CATATTATATATAATTATATTATATAATTATAGATAATACATATATTATCTATAATATAT 16847

Qy        479 GATATATGCTCACATTAGGTGTTTTGTTATGTATGGAGTGTTTTTCCCAAGGATATAAAA 538
              ||||||  |   |||  | |      || |||  | | | | |   |   ||||| ||
Db      16848 AATATATTATATATTTATATATAAATACATTTATATATTATATATGATAGATATATATAA 16907

Qy        539 GATTATTGATGCTTATTATTAGTTTTGTGCTATTTATAGCGATCTGCATGAGTCTTACTA 598
              ||||  |   | |   ||  | |   ||| || |   || || ||   | |   |
Db      16908 ATATATTTGTATATTTATATATGATAGATATATATAAATATATTTGTATATTTATATATG 16967

Qy        599 GAGGAGCAATACTTGTTGCTGCGCTTATTTACGCACTTTATATAATTTCAAATCGGAAGA 658
              ||  |||    |  |  | ||||||   |  || |||  |      |
Db      16968 ATAGATATATATAAATATATTTGTATATTTATATATGATAGATATATATTTGTATATTTA 17027

Qy        659 TGCTTTTTTGTGGAATAACTTTATTTGTAATAATTATACCCGTTTTAGCAATTTCTACTA 718
              |  |  | |   ||| | ||| | ||   ||||||     | ||  ||||| ||
Db      17028 TATATGATAGATATATATTTATATATTTATATATTATA-GATATATATAAATTTATATAT 17086

Qy        719 ATATTTTTGACAACTATACAGAAATTTTGATCGGCAGGTTTACAGATTCGTCTCAGGCAT 778
              ||| | | | |  |||| | |||| |   |    | | || |  | |    |  ||
Db      17087 TTATATATTATAGATATATATAAATATATTTATATATTTATATATTATAGATATATAAAT 17146

Qy        779 CGCGTGGATCTACACAGGGGCGGATAGATATGGCAATTAATTCATTAAACTTCCTGTCAG 838
              |           | | ||||   |||  ||| ||  |  |  | | | |
Db      17147 ATATTTATATATTTATATATTATAGATATATATAAATATATTTATATATTTATATATTAT 17206

Qy        839 AACATCCATCAGGTATAGGTCTGGGTACTCAAGGTTCAGGAAACATGCTTTCGGTAAAAG 898
              |         |  |||| | |   |    |   |  ||  |      | |   || |
Db      17207 AGATATATATAAATATATTTATATATTTATATATTATAGATATATAAATATATTTATATA 17266

Qy        899 ATAATAGGT--------TAAATACGGATAATTATTTTTTCTGGATCGCCCTTGAGACTGG 950
              | |||  |         || |||   |||  |||| | ||     |     |     |
Db      17267 TTTATATATTATAGATATATATAAATATATTTATATATTTATATTTATTGATATATATAA 17326

Qy        951 TATTATTGGCTTAATCATAAATATTATTTATCTGGCAAGTCAATTTTATTCTTCAACTTT 1010
                ||||    ||  | ||| ||  ||  |||  |   ||   | || ||   ||  |  ||
Db      17327 ATATATTTATATATTTATATATTATAGATATATATAAATATATTTATAGATTTATATATT 17386

Qy       1011 ACTAAATAGAATATATGGCAGTCATTGTAGCAATATGCACTATAGATTATATTTTCTCTT 1070
              |    |     ||| ||      |      |     ||||     ||| ||   | | | | |
Db      17387 ATAGATATATATAAATATATTTATAGATTTATATATTATAGATATATATAAATATATTTA 17446

Qy       1071 TGGAAGTATATATTTTATAAGTGCAGCGTTAAGTTCAGCACCTTCGTCATCAACTTTTTC 1130
              | ||  |||||||| || |  |  |    ||  || |     ||    || |    | |
Db      17447 TAGATTTATATATTATAGATATATATAAATATATTTATAGATTTATATATTATAGATATA 17506

Qy       1131 TATATATTATTGGACAGTTTTAGCTTTGATTCCATTTTTAAAATTAACAAATAGA 1185
              |||| ||   |  | || ||||  | | ||     | |  ||||  |   |||||
Db      17507 TATAAATATATTTATAGATTTATATATTATAGATATATATAAATATATTTATAGA 17561


RESULT 15
US-10-301-480-567933
; Sequence 567933, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 567933
;   LENGTH: 852
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-567933

  Query Match             3.8%;  Score 46;  DB 12;  Length 852;
  Best Local Similarity   54.8%;  Pred. No. 32;
  Matches   91;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy        925 TTTTTCTGGATCGCCCTTGAGACTGGTATTATTGGCTTAATCATAAATATTATTTATCTG 984
              | | |||   || |  |  | |    ||||||    |  || ||| ||||||| |
Db        423 TCTCTCTCTCTCTCTATATATATATATTATTATATTTATATTATATATATTATTATTATAT 482

Qy        985 GCAAGTCAATTTTATTCTTCAACTTTACTAAATAGAATATATGGCAGTCATTGTAGCAAT 1044
              |   |  | |||| ||    | || || |||  ||||||  | | ||  |  | | |
Db        483 ATATAATATATATATTATTATTATATATTATATAATATATATTATAATTATATAATTATT 542

Qy       1045 ATGCACTATAGATTATATTTTCTCTTTGGAAGTATATATTTTATAA 1090
              ||  |   || || |||| |  | | |   || |||||||| |||||
Db        543 ATATATATTATATAATATATAATATATATAATTATATATTATATAA 588


Search completed: May 12, 2006, 06:55:09
Job time : 615.399 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 08:31:04 ; Search time 4653.96 Seconds
(without alignments)
15792.702 Million cell updates/sec

Title: US-09-423-093B-4_COPY_12762_14054
Perfect score: 1293
Sequence: 1 gtgaaagttcaattgttaaa.....ttaaggcaaataagggataa 1293

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	1293	100.0	22043	1	AE008792	AE008792 Salmonell
	2	1293	100.0	22080	1	SERFBB	X56793 S.enterica
	3	1293	100.0	22080	6	BD085557	BD085557 Nucleic a
	4	1293	100.0	22080	6	AX029264	AX029264 Sequence
	5	537.8	41.6	2879	1	SERFBD1B	M65054 Salmonella
c	6	537.8	41.6	250050	1	AL627273	AL627273 Salmonell
	7	537.8	41.6	300928	1	AE016836	AE016836 Salmonell
	8	536.2	41.5	110000	1	CP000026_08	Continuation (9 of
	9	256.6	19.8	19457	1	AF461769	AF461769 Yersinia
	10	255	19.7	20125	1	AF461770	AF461770 Yersinia
	11	140	10.8	110000	1	BX950851_16	Continuation (17 o
	12	85.2	6.6	105320	2	AC116920	AC116920 Dictyoste
	13	85	6.6	94384	8	AC011718	AC011718 Homo sapi
	14	84.2	6.5	143331	8	AC091214	AC091214 Homo sapi
	15	83.8	6.5	170627	14	AC125567	AC125567 Rattus no
c	16	82.8	6.4	154625	14	CR936360	CR936360 Homo sapi
	17	82.4	6.4	349980	6	AX344555	AX344555 Sequence
	18	82	6.3	93791	14	AC138073	AC138073 Homo sapi
	19	81.2	6.3	7597	6	AX345915	AX345915 Sequence
c	20	80.4	6.2	76568	2	MBREV	AF538053 Monosiga
c	21	80	6.2	206462	14	CR753794	CR753794 Danio rer
	22	79.6	6.2	250743	2	AE014836	AE014836 Plasmodiu
c	23	79	6.1	67970	2	PFMAL1P3	AL031746 Plasmodiu
c	24	79	6.1	313050	2	PFA929352	AL929352 Plasmodiu
	25	78.6	6.1	341050	2	PFA929357	AL929357 Plasmodiu
c	26	78	6.0	212999	14	AC151201	AC151201 Bos tauru
c	27	77.4	6.0	170102	8	AC008079	AC008079 Homo sapi
c	28	77.2	6.0	162294	14	CR855307	CR855307 Danio rer
c	29	77.2	6.0	184838	8	AC090740	AC090740 Homo sapi
c	30	77	6.0	152209	8	HS1108D11	AL034419 Human DNA
	31	76.6	5.9	145992	8	AC134919	AC134919 Homo sapi
	32	76.4	5.9	114788	14	AC166330	AC166330 Glycine m
c	33	76.2	5.9	33701	14	AC119398	AC119398 Homo sapi
	34	76.2	5.9	159618	8	AC096750	AC096750 Homo sapi
	35	76.2	5.9	172816	8	AC093899	AC093899 Homo sapi
	36	76	5.9	217621	14	BX901924	BX901924 Danio rer
c	37	75.2	5.8	107252	14	CT009722	CT009722 Danio rer
c	38	74.6	5.8	162445	8	AL158151	AL158151 Human DNA
c	39	74.4	5.8	257109	2	AC116963	AC116963 Dictyoste
	40	74	5.7	110000	14	PFMAL8P1_03	Continuation (4 of
	41	73.8	5.7	842	6	CS083951	CS083951 Sequence
c	42	73.8	5.7	151391	8	HSJ730H16	AL122006 Human DNA
c	43	73.8	5.7	154561	5	CR391989	CR391989 Zebrafish
c	44	73.6	5.7	136713	8	AC112255	AC112255 Homo sapi
c	45	73.4	5.7	163907	14	CR548626	CR548626 Danio rer

ALIGNMENTS

RESULT 1
AE008792/c
LOCUS AE008792 22043 bp DNA linear BCT 09-AUG-2005
DEFINITION Salmonella typhimurium LT2, section 96 of 220 of the complete genome.
ACCESSION AE008792 AE006468
VERSION AE008792.1 GI:16420610
KEYWORDS .
SOURCE Salmonella typhimurium LT2
ORGANISM Salmonella typhimurium LT2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 22043)
AUTHORS McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W., Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F., Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A., Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W., Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
TITLE Complete genome sequence of Salmonella enterica serovar Typhimurium LT2
JOURNAL Nature 413 (6858), 852-856 (2001)
PUBMED 11677609
REFERENCE 2 (bases 1 to 22043)
AUTHORS .
CONSRTM The Salmonella typhimurium Genome Sequencing Project
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/, and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.PangeaSystems.com/ecocyc/

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB;
http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES Location/Qualifiers

source 1. .22043
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"

gene complement(71. .1064)
/gene="wzzB"
/note="synonym: STM2079"

CDS complement(71. .1054)
/gene="wzzB"
/note="chain length determinant protein (polysaccharide antigen chainregulator). (SW:WZZB_SALTY)"
/codon_start=1
/transl_table=11
/product="regulator of length of O-antigen component of lipopolysaccharide chains"
/protein_id="AAL20983.1"
/db_xref="GI:16420611"
/translation="MTVDSNTSSGRGNDPEQIDLIELLLQLWRGKMTIIVAVIIAILL
AVGYLMIAKEKWTSTAIITQPDAAQVATYTNALNVLYGGNAPKISEVQANFISRFSSA
FSALSEVLDNQKEREKLTIEQSVKGQALPLSVSYVSTTAEGAQRRLAEYIQQVDEEVA
KELEVDLKDNITLQTKTLQESLETQEVVAQEQKDLRIKQIEEALRYADEAKITQPQIQ
QTQDVTQDTMFLLGSDALKSMIQNEATRPLVFSPAYYQTKQTLLDIKNLKVTADTVHV
YRYVMKPTLPVRRDSPKTAITLVLAVLLGGMIGAGIVLGRNALRSYKPKAL"

RBS complement(1059. .1064)
/gene="wzzB"
/note="putative RBS for wzzB; RegulonDB:STMS1H002228"

gene complement(1197. .2363)
/gene="udg"
/note="synonym: STM2080"

CDS complement(1197. .2363)
/gene="udg"
/EC_number="1.1.1.22"
/note="UDP-glucose 6-dehydrogenase. (SW:UDG_SALTY)"
/codon_start=1
/transl_table=11
/product="UDP-glucose/GDP-mannose dehydrogenase"
/protein_id="AAL20984.1"
/db_xref="GI:16420612"
/translation="MKITISGTGYVGLSNGLLIAQHHDVVALDIVPSRVELLNDRISP
IVDKEIQQFLKEDNIRFRATLDKFDAYQNADYVIIATPTDYDPKTNYFNTSSVESVMQ
DVISINPAAVMIIKSTVPVGFTAAMRQKFATENIIFSPEFLREGKALYDNLYPSRIVI
GEQSERAREFAALLQEGAIKQEIPTLFTDSTEAEAIKLFANTYLAMRVAYFNELDSYA
ETLGLNTRQIIEGVCLDPRIGNHYNNPSFGYGGYCLPKDTKQLLANYQSVPNNIISAI
VEANRTRKDFIADAILARKPKVVGIYRLIMKSGSDNFRASSIQGIMKRIKAKGVEVII
YEPVMEEDTFFNSRLERDLHCFKQQADVIISNRMAAELLDVAEKVYTRDLFGSD"

gene complement(2600. .4018)
/gene="gnd"
/note="synonym: STM2081"

CDS complement(2600. .4006)
/gene="gnd"
/EC_number="1.1.1.44"
/note="decarboxylating; 6-phosphogluconate dehydrogenase, decarboxylating. (SW:6PGD_SALTY)"
/codon_start=1
/transl_table=11
/product="gluconate-6-phosphate dehydrogenase"
/protein_id="AAL20985.1"
/db_xref="GI:16420613"
/translation="MSKQQIGVVGMAVMGRNLALNIESRGYTVSVFNRSREKTEEVIA
ENPGKKLVPYYTVKEFVESLETPRRILLMVKAGAGTDAAIDSLKPYLEKGDIIIDGGN
TFFQDTIRRNRELSAEGFNFIGTGVSGGEEGALKGPSIMPGGQKDAYELVAPILTKIA
AVAEDGEPCVTYIGADGAGHYVKMVHNGIEYGDMQLIAEAYSLLKGGLNLSNEELANT
FTEWNNGELSSYLIDITKDIFTKKDEDGNYLVDVILDEAANKGTGKWTSQSALDLGEP
LSLITESVFARYISSLKAQRVAASKVLSGPKAQPAGDKAEFIEKVRRALYLGKIVSYA
QGFSQLRAASDEYHWDLNYGEIAKIFRAGCIIRAQFLQKITDAYAENADIANLLLAPY
FKKIADEYQQALRDVVAYAVQNGIPVPTFSAAVAYYDSYRAAVLPANLIQAQRDYFGA
HTYKRTDKEGIFHTEWLE"

RBS complement(4013. .4018)
/gene="gnd"
/note="putative RBS for gnd; RegulonDB:STMS1H002229"

gene complement(4170. .5607)
/gene="rfbP"
/note="synonym: STM2082"

CDS complement(4170. .5600)
/gene="rfbP"
/note="bifunctional enzyme; LPS side chain defect; undecaprenyl-phosphate galactosephosphotransferase. (SW:RFBP_SALTY); O-antigen transfer protein"
/codon_start=1
/transl_table=11
/product="undecaprenol-phosphate galactosephosphotransferase, and"
/protein_id="AAL20986.1"
/db_xref="GI:16420614"
/translation="MDNIDNKYNPQLCKIFLAISDLIFFNLALWFSLGCVYFIFDQVQ
RFIPQDQLDTRVITHFILSVVCVGWFWIRLRHYTYRKPFWYELKEIFRTIVIFAIFDL
ALIAFTKWQFSRYVWVFCWTFALILVPFFRALTKHLLNKLGIWKKKTIILGSGQNARG
AYSALQSEEMMGFDVIAFFDTDASDAEINMLPVIKDTEIIWDLNRTGDVHYILAYEYT
ELEKTHFWLRELSKHHCRSVTVVPSFRGLPLYNTDMSFIFSHEVMLLRIQNNLAKRSS
RFLKRTFDIVCSIMILIIASPLMIYLWYKVTRDGGPAIYGHQRVGRHGKLFPCYKFRS
MVMNSQEVLKELLANDPIARAEWEKDFKLKNDPRITAVGRFIRKTSLDELPQLFNVLK
GDMSLVGPRPIVSDELERYCDDVDYYLMAKPGMTGLWQVSGRNDVDYDTRVYFDSWYV
KNWTLWNDIAILFKTAKVVLRRDGAY"

RBS complement(5602. .5607)
/gene="rfbP"
/note="putative RBS for rfbP; RegulonDB:STMS1H002230"

gene complement(5672. .7112)
/gene="rfbK"
/note="synonym: STM2083"

CDS complement(5672. .7105)
/gene="rfbK"
/EC_number="5.4.2.8"
/EC_number="1.1.1.44"
/note="LPS side chain defect: phosphomannomutase. (SW:RFBK_SALTY)"
/codon_start=1
/transl_table=11
/product="phosphomannomutase"
/protein_id="AAL20987.1"
/db_xref="GI:16420615"
/translation="MNVVNNSRDVIYSSGIVFGTSGARGLVKDFTPQVCAAFTVSFVA
VMQEHFSFDTVALAIDNRPSSYGMAQACAAALADKGVNCIFYGVVPTPALAFQSMSDN
MPAIMVTGSHIPFERNGLKFYRPDGEITKHDEAAILSVEDTCSHLELKELIVSEMAAV
NYISRYTSLFSTPFLKNKRIGIYEHSSAGRDLYKPLFIALGAEVVSLGRSDNFVPIDT
EAVSKEDREKARSWAKEFDLDAIFSTDGDGDRPLIADEAGEWLRGDILGLLCSLALDA
EAVAIPVSCNSIISSGRFFKHVKLTKIGSPYVIEAFNELSRSYSRIVGFEANGGFLLG
SDICINEQNLHALPTRDAVLPAIMLLYKSRNTSISALVNELPTRYTHSDRLQGITTDK
SQSLISMGRENLSNLLSYIGLENEGAISTDMTDGMRITLRDGCIVHLRASGNAPELRC
YAEANLLNRAQDLVNTTLANIKKRCLL"

gene complement(7092. .8534)
/gene="rfbM"
/note="synonym: STM2084"

CDS complement(7092. .8531)
/gene="rfbM"
/EC_number="2.7.7.22"
/EC_number="1.1.1.44"
/note="LPS side chain defect: mannose-1-phosphate guanylyltransferase GDP). (SW:RFBM_SALTY)"
/codon_start=1
/transl_table=11
/product="mannose-1-phosphate guanylyltransferase"

```
                     /protein_id="AAL20988.1"
                     /db_xref="GI:16420616"
                     /translation="MSFLPVIMAGGTGSRLWPLSREYHPKQFLSVEGKLSMLQNTIKR
                     LASLSTEEPVVICNDRHRFLVAEQLREIDKLANNIILEPVGRNTAPAIALAAFCALQN
                     ADNADPLLLVLAADHVIQDEIAFTKAVRHAEEYAANGKLVTFGIVPTHAETGYGYIRR
                     GELIGNDAYAVAEFVEKPDIDTAGDYFKSGKYYWNSGMFLFRASSYLNELKYLSPEIY
                     KACEKAVGHINPDLDFIRIDKEEFMSCPSDSIDYAVMEHTQHAVVIPMSAGWSDVGSW
                     SSLWDISNKDHQRNVLKGDIFAHACNDNYIYSEDMFISAIGVSNLVIVQTTDALLVAN
                     KDTVQDVKKIVDYLKRNDRNEYKQHQEVFRPWGKYNVIDSGKNYLVRCITVKPGEKFV
                     AQMHHHRAEHWIVLSGTARVTKGEQTYMVSENESTFIPPNTIHALENPGMTPLKLIEI
                     QSGTYLGEDDIIRLEQRSGFSKEWTNERS"
     RBS             complement(7107. .7112)
                     /gene="rfbK"
                     /note="putative RBS for rfbK; RegulonDB:STMS1H002231"
     RBS             complement(8529. .8534)

  Query Match           100.0%;  Score 1293;  DB 1;  Length 22043;
  Best Local Similarity  100.0%;  Pred. No. 3.8e-187;
  Matches 1293;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13156 GTGAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTA 13097

Qy         61 TCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13096 TCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTA 13037

Qy        121 GGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13036 GGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCT 12977

Qy        181 GTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12976 GTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAAC 12917

Qy        241 AAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12916 AAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTT 12857

Qy        301 TTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12856 TTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTT 12797

Qy        361 CATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12796 CATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGT 12737

Qy        421 TCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12736 TCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAA 12677

Qy        481 GCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12676 GCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATAC 12617

Qy        541 TATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12616 TATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGT 12557

Qy        601 ATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12556 ATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACA 12497

Qy        661 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12496 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 12437

Qy        721 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12436 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 12377

Qy        781 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12376 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 12317

Qy        841 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12316 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 12257

Qy        901 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12256 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 12197

Qy        961 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12196 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 12137

Qy       1021 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12136 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 12077

Qy       1081 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12076 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 12017

Qy       1141 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12016 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 11957

Qy       1201 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11956 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCA 11897

Qy       1261 CTAACTTACTTAATTAAGGCAAATAAGGGATAA 1293
              |||||||||||||||||||||||||||||||||
Db      11896 CTAACTTACTTAATTAAGGCAAATAAGGGATAA 11864
```

```
RESULT 2
SERFBB
LOCUS       SERFBB                 22080 bp    DNA     linear   BCT 14-JUN-2005
DEFINITION  S.enterica (group B) rfb gene cluster.
ACCESSION   X56793 M29713
VERSION     X56793.1  GI:47885
KEYWORDS    abe gene; CDP-abequose synthase; CDP-glucose 4.6-dehydratase; ddhA
            gene; ddhB gene; ddhC gene; ddhD gene; dTDP-6-deoxy-L-mannose
            dehydrogenase; dTDP-dehydrorhamnose reductase; dTDP-glucose
            oxidoreductase; dTDP-glucose-4.6-dehydratase; galF gene;
            glucose-1-P cytidyltransferase; glucose-1-P thymidyltransferase;
            manB gene; manC gene; Manose-1-Pguanylyltransferase;
            oxidoreductase; Phosphomanomutase; rfb gene cluster; rmlA gene;
            rmlB gene; rmlC gene; rmlD gene; undecaprenyl-P galactose-P
            transferase; wbaN gene; wbaP gene; wbaU gene; wbaV gene; wcaL gene;
            wzx gene.
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22080)
  AUTHORS   Wyk,P. and Reeves,P.
  TITLE     Identification and sequence of the gene for abequose synthase,
            which confers antigenic specificity on group B salmonellae:
            homology with galactose epimerase
  JOURNAL   J. Bacteriol. 171 (10), 5687-5693 (1989)
   PUBMED   2793832
REFERENCE   2  (bases 1 to 22080)
  AUTHORS   Jiang,X.M., Neal,B., Santiago,F., Lee,S.J., Romana,L.K. and
            Reeves,P.R.
  TITLE     Structure and sequence of the rfb (O antigen) gene cluster of
            Salmonella serovar typhimurium (strain LT2)
  JOURNAL   Mol. Microbiol. 5 (3), 695-713 (1991)
   PUBMED   1710759
REFERENCE   3  (bases 1 to 22080)
  AUTHORS   Reeves,P.
  TITLE     Evolution of Salmonella O antigen variation by interspecific gene
```

```
            transfer on a large scale
  JOURNAL   Trends Genet. 9 (1), 17-22 (1993)
   PUBMED   8434412
REFERENCE   4  (bases 1 to 22080)
  AUTHORS   Liu,D., Haase,A.M., Lindqvist,L., Lindberg,A.A. and Reeves,P.R.
  TITLE     Glycosyl transferases of O-antigen biosynthesis in Salmonella
            enterica: identification and characterization of transferase genes
            of groups B, C2, and E1
  JOURNAL   J. Bacteriol. 175 (11), 3408-3413 (1993)
   PUBMED   7684736
REFERENCE   5  (bases 1 to 22080)
  AUTHORS   Stevenson,G., Andrianopoulos,K., Hobbs,M. and Reeves,P.R.
  TITLE     Organization of the Escherichia coli K-12 gene cluster responsible
            for production of the extracellular polysaccharide colanic acid
  JOURNAL   J. Bacteriol. 178 (16), 4885-4893 (1996)
   PUBMED   8759852
REFERENCE   6  (bases 1 to 22080)
  AUTHORS   Marolda,C.L. and Valvano,M.A.
  TITLE     The GalF protein of Escherichia coli is not a UDP-glucose
            pyrophosphorylase but interacts with the GalU protein possibly to
            regulate cellular levels of UDP-glucose
  JOURNAL   Mol. Microbiol. 22 (5), 827-840 (1996)
   PUBMED   8971705
REFERENCE   7  (bases 1 to 22080)
  AUTHORS   Reeves,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-1990) P.R. Reeves, G08 DEPT OF MICROBIOLOGY,
            UNIVERSITY OF SYDNEY, NSW 2006, AUSTRALIA
COMMENT     On Jun 13, 2005 this sequence version replaced gi:154337.
FEATURES             Location/Qualifiers
     source          1. .22080
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /strain="group B, serovar typhimurium, strain LT2"
                     /isolate="SL1654"
                     /db_xref="taxon:28901"
     gene            17. .1237
                     /gene="wcaL"
     CDS             17. .1237
                     /gene="wcaL"
                     /note="part of cluster for colanic acid prodaction"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA40132.1"
                     /db_xref="GI:47886"
                     /db_xref="GOA:P26388"
                     /db_xref="InterPro:IPR001296"
                     /db_xref="UniProt/Swiss-Prot:P26388"
                     /translation="MKVSFFLLKFPLSSETFVLNQITAFIDMGHEVEIVALQKGDTQH
                     THAAWEKYGLAAKTRWLQDEPQGRLAKLRYRACKTLPGLHRAATWKALNFTRYGDESR
                     NLILSAICAQVSQPFVADVFIAHFGPAGVTAAKLRELGVLRGKIATIFHGIDISSREV
                     LSHYTPEYQQLFRRGDLMLPISDLWAGRLKSMGCPPEKIAVSRMGVDMTRFTHRSVKA
                     PGMPLEMISVARLTEKKGLHVAIEACRQLKAQGVAFRYRILGIGPWERRLRTLIEQYQ
                     LEDVIEMPGFKPSHEVKAMLDDADVFLLPSITGTDGDMEGIPVALMEAMAVGIPVVST
                     VHSGIPELVEAGKSGWLVPENDAQALAARLAEFSRIDHDTLESVITRAREKVAQDFNQ
                     QAINRQLASLLQTI"
     gene            1248. .2651
                     /gene="ORF 1.3"
     CDS             1248. .2651
                     /gene="ORF 1.3"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA40113.1"
                     /db_xref="GI:47887"
                     /db_xref="GOA:P26389"
                     /db_xref="UniProt/Swiss-Prot:P26389"
                     /translation="MPATKFSRRTLLTAGSALAVLPFLRALPVQAREPRETVDIKDYP
                     ADDGIASFKQAFADGQTVVVPPGWVCENINAAITIPAGKTLRVQGAVRGNGRGRFILQ
                     DGCQVVGEQGGSLHNVTLDVRGSDCVIKGVAMSGFGPVAQIFIGGKEPQVMRNLIIDD
                     ITVTHANYAILRQGFHNQMDGARITHSRFSDLQGDAIEWNVAIHDRDILISDHVIERI
                     NCTNGKINWGIGIGLAGSTYDNSYPEDQAVKNFVVANITGSDCRQLVHVENGKHFVIR
                     NVKAKNITPGFSKNAGIDNATIAIYGCDNFVIDNIDMTNSAGMLIGYGVVKGKYLSIP
                     QNFKLNAIRLDNRQVAYKLRGIQISSGNTPSFVAITNVRMTRATLELHNQPQHLFLRN
                     INVMQTSAIGPALKMHFDLRKDVRGQFMARQDTLLSLANVHAINENGQSSVDIDRINH
                     QTVNVEAVNFSLPKRGG"
     gene            2829. .3722
                     /gene="galF"
     CDS             2829. .3722
                     /gene="galF"
                     /function="possibly in regulation of cellular levels of
                     UDP-glucose"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glucose-1-phosphate
                     uridylyltransferase"
                     /protein_id="CAA40114.1"
                     /db_xref="GI:47888"
                     /db_xref="GOA:P0A2K7"
                     /db_xref="InterPro:IPR005774"
                     /db_xref="InterPro:IPR005835"
                     /db_xref="UniProt/Swiss-Prot:P0A2K7"
                     /translation="MMNLKAVIPVAGLGMHMLPATKAIPKEMLPIVDKPMIQYIVDEI
                     VAAGIKEIVLVTHASKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNV
                     RQAQPLGLGHSILCARPVVGDNPFIVVLPDIIIDDATADPLRYNLAAMVARFNETGRS
                     QVLAKRMKGDLSEYSVIQTKEPLDNEGKVSRIVEFIEKPDQPQTLDSDLMAVGRYVLS
                     ADIWAELERTEPGAWGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKY
                     GLRNLKEGAKFRKSIEQLLHE"
     gene            4099. .5184
                     /gene="rmlB"
     CDS             4099. .5184
                     /gene="rmlB"
                     /EC_number="4.2.1.46"
                     /note="CDS starts with gtg"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="dTDP-glucose 4,6-dehydratase"
                     /protein_id="CAA40115.1"
                     /db_xref="GI:581654"
                     /db_xref="GOA:P26391"
                     /db_xref="InterPro:IPR001509"
                     /db_xref="InterPro:IPR005888"
                     /db_xref="UniProt/Swiss-Prot:P26391"
                     /translation="MKILITGGAGFIGSAVVRHIIKNTQDTVVNIDKLTYAGNLESLS
                     DISESNRYNFEHADICDSAEITRIFEQYQPDAVMHLAAESHVDRSITGPAAFIETNIV
                     GTYALLEVARKYWSALGEDKKNNFRFHHISTDEVYGDLPHPDEVENSVTLPLFTETTA
                     YAPSSPYSASKASSDHLVRAWRRTYGLPTIVTNCSNNYGPYHFPEKLIPLVILNALEG
                     KPLPIYGKGDQIRDWLYVEDHARALHMVVTEGKAGETYNIGGHNEKKNLDVVFTICDL
                     LDEIVPKATSYREQITYVADRPGHDRRYAIDAGKISRELGWKPLETFESGIRKTVEWY
                     LANTQWVNNVKSGAYQSWIEQNYEGRQ"
     gene            5184. .6083
                     /gene="rmlD"
     CDS             5184. .6083
                     /gene="rmlD"
                     /note="incorrectly designated rfbA in reference [2];
                     dTDP-glucose oxidoreductase"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="dTDP-6-deoxy-L-mannose dehydrogenase"
                     /protein_id="CAA40116.1"
                     /db_xref="GI:47890"
                     /db_xref="GOA:P26392"
                     /db_xref="InterPro:IPR005913"
                     /db_xref="UniProt/Swiss-Prot:P26392"
                     /translation="MNILLFGKTGQVGWELQRSLAPVGNLIALDVHSKEFCGDFSNPK
                     GVAETVRKLRPDVIVNAAAHTAVDKAESEPELAQLLNATSVEAIAKAANETGAWVVHY
                     STDYVFPGTGDIPWQETDATSPLNVYGKTKLAGEKALQDNCPKHLIFRTSWVYAGKGN
                     NFAKTMLRLAKERQTLSVINDQYGAPTGAELLADCTAHAIRVALNKPEVAGLYHLVAG
                     GTTTWHDYAALVFDEARKAGITLALTELNAVPTSAYPTPASRPGNSRLNTEKFQRNFD
                     LILPQWELGVKRMLTEMFTTTTI"
     gene            6131. .7009
                     /gene="rmlA"
     CDS             6131. .7009
                     /gene="rmlA"
                     /note="referred to as orf6.1 in reference [2]"
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="glucose-1-phosphate thymidylyltransferase"
                     /protein_id="CAA40117.1"
                     /db_xref="GI:47891"
                     /db_xref="GOA:P26393"
                     /db_xref="InterPro:IPR005835"
                     /db_xref="InterPro:IPR005907"
                     /db_xref="UniProt/Swiss-Prot:P26393"
                     /translation="MKTRKGIILAGGSGTRLYPVTMAVSKQLLPIYDKPMIYYPLSTL
                     MLAGIRDILIISTPQDTPRFQQLLGDGSQWGLNLQYKVQPSPDGLAQAFIIGEEFIGH
                     DDCALVLGDNIFYGHDLPKLMEAAVNKESGATVFAYHVNDPERYGVVEFDQKGTAVSL
                     EEKPLQPKSNYAVTGLYFYDNSVVEMAKNLKPSARGELEITDINRIYMEQGRLSVAMM
                     GRGYAWLDTGTHQSLIEASNFIATIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPL
                     SKNDYGKYLLKMVKGL"
     gene            7010. .7561
                     /gene="rmlC"
     CDS             7010. .7561
                     /gene="rmlC"

  Query Match             100.0%;  Score 1293;  DB 1;  Length 22080;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-187;
  Matches 1293;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12762 GTGAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTA 12821

Qy         61 TCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12822 TCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTA 12881

Qy        121 GGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12882 GGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCT 12941

Qy        181 GTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12942 GTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAAC 13001

Qy        241 AAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13002 AAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTT 13061

Qy        301 TTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13062 TTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTT 13121

Qy        361 CATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13122 CATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGT 13181

Qy        421 TCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13182 TCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAA 13241

Qy        481 GCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13242 GCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATAC 13301

Qy        541 TATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13302 TATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGT 13361

Qy        601 ATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13362 ATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACA 13421

Qy        661 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13422 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 13481

Qy        721 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13482 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 13541

Qy        781 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13542 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 13601

Qy        841 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13602 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 13661

Qy        901 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13662 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 13721

Qy        961 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13722 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 13781

Qy       1021 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13782 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 13841

Qy       1081 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13842 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 13901

Qy       1141 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13902 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 13961

Qy       1201 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13962 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCA 14021

Qy       1261 CTAACTTACTTAATTAAGGCAAATAAGGGATAA 1293
              ||||||||||||||||||||||||||||||||
Db      14022 CTAACTTACTTAATTAAGGCAAATAAGGGATAA 14054


RESULT 3
BD085557
LOCUS       BD085557               22080 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085557
VERSION     BD085557.1  GI:22631167
KEYWORDS    JP 2001522248-A/4.
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22080)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 4 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Salmonella enterica
            PN   JP 2001522248-A/4
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .22080
            FT                   /organism='Salmonella enterica'.
FEATURES             Location/Qualifiers
     source          1. .22080
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:28901"
ORIGIN

  Query Match           100.0%;  Score 1293;  DB 6;  Length 22080;
  Best Local Similarity  100.0%;  Pred. No. 3.8e-187;
  Matches 1293;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12762 GTGAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTA 12821

Qy         61 TCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12822 TCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTA 12881

Qy        121 GGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12882 GGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCT 12941

Qy        181 GTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12942 GTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAAC 13001

Qy        241 AAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13002 AAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTT 13061

Qy        301 TTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13062 TTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTT 13121

Qy        361 CATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13122 CATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGT 13181

Qy        421 TCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13182 TCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAA 13241

Qy        481 GCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13242 GCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATAC 13301

Qy        541 TATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13302 TATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGT 13361

Qy        601 ATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13362 ATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACA 13421

Qy        661 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13422 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 13481

Qy        721 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13482 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 13541

Qy        781 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13542 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 13601

Qy        841 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13602 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 13661

Qy        901 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13662 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 13721

Qy        961 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      13722 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 13781

Qy       1021 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13782 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 13841

Qy       1081 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13842 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 13901

Qy       1141 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13902 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 13961

Qy       1201 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13962 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCA 14021

Qy       1261 CTAACTTACTTAATTAAGGCAAATAAGGGATAA 1293
              |||||||||||||||||||||||||||||||||
Db      14022 CTAACTTACTTAATTAAGGCAAATAAGGGATAA 14054


RESULT 4
AX029264
LOCUS       AX029264               22080 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 4 from Patent WO9850531.
ACCESSION   AX029264
VERSION     AX029264.1  GI:10190115
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 4 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .22080
                     /organism="Salmonella enterica"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match           100.0%;  Score 1293;  DB 6;  Length 22080;
  Best Local Similarity  100.0%;  Pred. No. 3.8e-187;
  Matches 1293;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12762 GTGAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTA 12821

Qy         61 TCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12822 TCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTA 12881

Qy        121 GGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12882 GGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCT 12941

Qy        181 GTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12942 GTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAAC 13001

Qy        241 AAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13002 AAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTT 13061

Qy        301 TTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      13062 TTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTT 13121

Qy        361 CATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13122 CATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGT 13181

Qy        421 TCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13182 TCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAA 13241

Qy        481 GCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13242 GCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATAC 13301

Qy        541 TATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13302 TATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGT 13361

Qy        601 ATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13362 ATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACA 13421

Qy        661 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13422 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 13481

Qy        721 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13482 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 13541

Qy        781 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13542 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 13601

Qy        841 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13602 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 13661

Qy        901 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13662 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 13721

Qy        961 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13722 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 13781

Qy       1021 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13782 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 13841

Qy       1081 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13842 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 13901

Qy       1141 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13902 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 13961

Qy       1201 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13962 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCA 14021

Qy       1261 CTAACTTACTTAATTAAGGCAAATAAGGGATAA 1293
              |||||||||||||||||||||||||||||||
Db      14022 CTAACTTACTTAATTAAGGCAAATAAGGGATAA 14054


RESULT 5
SERFBD1B
LOCUS       SERFBD1B                2879 bp    DNA     linear   BCT 12-MAR-1997
DEFINITION  Salmonella enterica group D1 wzx gene, partial cds and wbaV gene,
            complete cds.
ACCESSION   M65054
VERSION     M65054.1  GI:154334
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2879)
  AUTHORS   Liu,D., Verma,N.K., Romana,L.K. and Reeves,P.R.
  TITLE     Relationships among the rfb regions of Salmonella serovars A, B,
            and D
  JOURNAL   J. Bacteriol. 173 (15), 4814-4819 (1991)
   PUBMED   1856174
REFERENCE   2  (bases 1 to 2879)
  AUTHORS   Reeves,P.
  TITLE     Evolution of Salmonella O antigen variation by interspecific gene
            transfer on a large scale
  JOURNAL   Trends Genet. 9 (1), 17-22 (1993)
   PUBMED   8434412
REFERENCE   3  (bases 1 to 2879)
  AUTHORS   Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C.,
            Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
  TITLE     Bacterial polysaccharide synthesis and gene nomenclature
  JOURNAL   Trends Microbiol. 4 (12), 495-503 (1996)
   PUBMED   9004408
FEATURES             Location/Qualifiers
     source          1. .2879
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /strain="Ty2, group D1, serovar typhi"
                     /db_xref="taxon:28901"
                     /note="polysaccharide genes have been named in accordance
                     with the Bacterial Polysaccharide Gene Nomenclature
                     scheme"
     gene            1. .1299
                     /gene="wzx"
     CDS             <1. .1299
                     /gene="wzx"
                     /function="putative O-antigen exporter"
                     /note="previously called rfbX; referred to as orf12.8 in
                     citation [2]; bases 1-28 overlaps with bases 2283-2310 of
                     M29682"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="putative 12 transmembrane segment protein"
                     /protein_id="AAB49394.1"
                     /db_xref="GI:154335"
                     /translation="VRKLRLVRIPRHLIIAASSWLSKIIIAGVQLVSVKFLLEILGEE
                     SYAVFTLLTGLLVWFSIADIGIGSSLQNYISELKADRKSYDAYIKAAVHILFASLIIL
                     SSTLFFLSDKLSSLYLTSFSDELKNNSGSYFFIASILFIFIGVGSVVYKILFAELLGW
                     KANIINALSYLLGFLDVVAIHYLMPDSSITFALVALYAPVAILPIIYISFRYIYVLKA
                     KVNFNTYKLLLSRSSGFLIFSSLSIIVLQTDYIVMSQKLSAADIIKYTVTMKIFGLMF
                     FIYTAVLQALWPVCAELRVKMQWRKLHRIIFLNIIGGVFFIGLGTLFIYVLKDYIYSI
                     IANGIDYNISGVVFVLLAVYFSIRVWCDTFAMLLQSMNQLKILWLIVPCQALIGGVTQ
                     WYFAEHYGIVGILYGLILSFSLTVFWGLPVYYMYKSKRLA"
     gene            1301. .2302
                     /gene="wbaV"
     CDS             1301. .2302
                     /gene="wbaV"
                     /note="orf14.1; possibly encodes abequosyl transferase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB49395.1"
                     /db_xref="GI:154336"
                     /translation="MKVSFCIPTYNRVKFIEDLLESINNQSSHSLIVEVCISDNASTD
                     GTEESINIWRDRFNFPILYQRHNENIGPDRNYLSAVNMGTGDYCWIFGSDDILTKNSL
                     ALMEDKLAAGSDIYLCDRRELDISMTKISNPHRRWLNGGSRLFSFSNEADLIEYFSKC
                     NSVGGLFSYLSSIIVKRNKWSDVLFDESYIGTAYAHVYILLRIINNMNSTLQYISLPL
                     VDCRGDNDTFESNGKARRIKIDFIGYLKLREDFYNNNTKIYISFGRVLTKERPWLYTS
                     LAMACYGDSTDRAELASFYKKLGYPKIATNLIFCLKGLASYTKKIKLAKMVIKKIFS"
ORIGIN

  Query Match             41.6%;  Score 537.8;  DB 1;  Length 2879;
```

```
  Best Local Similarity   63.7%;  Pred. No. 2.5e-72;
  Matches  818;  Conservative    0;  Mismatches  467;  Indels    0;  Gaps    0;

Qy          3 GAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTATC 62
              ||||  |     ||| | | |||||| || ||| | ||| |||| | ||| || ||| | ||
Db          6 GAAACTGAGGTTGGTTAGAATTCCAAGACATCTTATTATTGCCGCTTCCTCTTGGCTTTC 65

Qy         63 CAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTAGG 122
              || ||||||||||||||| || ||||||| |||| ||  || |||| |     || || ||
Db         66 AAAGATAATAATTGCCGGTGTTCAGTTAGTAAGTGTTAAATTTCTTTTAGAAATTCTTGG 125

Qy        123 TGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCTGT 182
              |||||   ||| ||  | |||| | | |||||||| ||||| || |||| ||||  ||
Db        126 CGAAGAATCATACGCTGTATTTACTCTTTTAACTGGATTATTGGTCTGGTTTAGCATTGC 185

Qy        183 TGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAACAA 242
              ||| |||| || ||||    | || ||||||||||||||| || |  | |||  | |  ||
Db        186 AGATATTGGGATTGGTAGTAGTCTACAAAATTATATATCTGAGTTGAAAGCTGATAGAAA 245

Qy        243 AAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTT 302
              |   |||||||||||||| ||  | ||| | ||| |    ||   | |  | ||  ||||
Db        246 ATCATATGATGCATATATCAAGGCCGCAGTTCATATTCTATTCGCATCCTTAATCATTTT 305

Qy        303 TATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCA 362
              |   ||  ||| |   | || || |   | | ||  |  |||||| |||| |||
Db        306 AAGCTCTACATTATTCTTCTTATCAGATAAATTATCGTCACTATATCTTACTTCATTTAG 365

Qy        363 TGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTC 422
              || | |||  |  |||  |  |||      |||||||   || || |  | || |  |
Db        366 CGATGAATTGAAAAACAACTCAGGAAGTTATTTTTTTATAGCAAGTATATTATTTATATT 425

Qy        423 TATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGC 482
              || ||  | ||   | | |  |||||||| | ||||| ||| || | ||||||||||
Db        426 CATCGGCGTTGGGAGTGTGGTCTATAAAATATTATTTGCGGAACTGTTAGGGTGGAAAGC 485

Qy        483 TAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTA 542
              |||| || | || ||||||||||| |  ||||| |  | |   |  |    ||  | ||
Db        486 TAATATAATTAATGCATTATCTTATCTTTTAGGTTTTTTAGATGTAGTTGCGATCCATTA 545

Qy        543 TAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTAT 602
              |     |              || |  ||  |  || |||   |||||  ||| || |  ||
Db        546 TTTAATGCCAGATTCGAGTATTACCTTCGCTTTAGTAGCATTGTATGCTCCGGTAGCAAT 605

Qy        603 GATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAA 662
              |  |  |    |||||    | | | || ||| |  ||  | | |  |||  ||
Db        606 ACTGCCCATTATATATATATCGTTTCGGTATATATATGTTCTTAAAGCGAAAGTAAACTT 665

Qy        663 ATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATC 722
                  |||| |  |  ||    | || |||||||| |  ||||| | |  |||||
Db        666 TAACACCTATAAATTATTACTATCACGTTCATCAGGGTTTCTGATTTTTTCGTCCTTATC 725

Qy        723 GATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATAT 782
              |||| | ||  | ||||| |||||||| || || |||||  |   || || |||| |||||
Db        726 GATAATAGTTTTACAAACTGATTATATTGTGATGTCTCAGAAATTATCTGCTGCAGATAT 785

Qy        783 TGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTAT 842
              | |  ||||||| |||||||||||||| ||||||||| | ||||||||||||||||  |
Db        786 TATAAAATATACTGTAACGATGAAAATATTTGGTTTAATGTTTTTTATTTATACTGCGGT 845

Qy        843 TTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACT 902
              || ||||||||||||| ||||||||||||  |||| |||  ||| |||| ||| ||
Db        846 ATTACAAGCATTATGGCCAGTATGTGCTGAATTACGAGTGAAAATGCAGTGGAGAAAGCT 905

Qy        903 TAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAAT 962
              | | ||| ||     | |||||| |   |||   | | | | ||| |    ||||  |
Db        906 GCATAGAATCATTTTCCTAAATATTATTGGTGGGGTATTTTTTATTGGTCTTGGTACGTT 965

Qy        963 ATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCA 1022
              |||||||||| | || || ||  | || | |    |||| || || | |||  ||||| |
Db        966 ATTTATTTATGTTTTAAAGGATTATATCTATAGCATAATTGCTAACGGTATAGATTATAA 1025

Qy       1023 AGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGA 1082
                ||  |    |  |||| | ||| | |   | |||||  ||||  | ||||||||||
```

```
Db       1026 TATTAGTGGGGTTGTTTTTGTTTTACTGGCTGTGTATTTTAGTATAAGAGTTTGGTGTGA 1085

Qy       1083 CACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTACC 1142
              || | ||| |||||| | |||||||||||  | |||||||| ||||||  |  |||| ||
Db       1086 TACATTTGCTATGTTACTTCAAAGTATGAACCAATTAAAAATTCTTTGGCTCATAGTTCC 1145

Qy       1143 ACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATCAG 1202
                 || ||| ||||||||||  |   | |||||||||||| |          ||||||
Db       1146 GTGTCAGGCATTAATTGGTGGTGTGACTCAATGGTATTTTGCAGAGCATTATGGAATAGT 1205

Qy       1203 TGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCACT 1262
              |||  |  |   ||  | ||| |||| ||  |  |||||||||||||||  | ||| |
Db       1206 TGGTATTTTATACGGACTAATTTTATCGTTCTCGCTAACTGTTTTTTGGGGATTGCCAGT 1265

Qy       1263 AACTTACTTAATTAAGGCAAATAAG 1287
                 |||  |   ||||   || | |
Db       1266 GTATTATATGTATAAGAGTAAAAGG 1290
```

RESULT 6
AL627273/c

```
LOCUS       AL627273               250050 bp    DNA     linear   BCT 16-APR-2005
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 9/20.
ACCESSION   AL627273 AL513382
VERSION     AL627273.1  GI:16503247
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 250050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
   PUBMED   11677608
REFERENCE   2  (bases 1 to 250050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .250050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            complement(109. .1566)
                     /gene="STY2216"
     CDS             complement(109. .1566)
                     /gene="STY2216"
                     /note="Orthologue of E. coli YEEO_ECOLI; Fasta hit to
                     YEEO_ECOLI (547 aa), 81% identity in 488 aa overlap
                     Unknown function. Contains hydrophobic, possible
                     membrane-spanning regions"
                     /codon_start=1
                     /transl_table=11
                     /product="putative inner membrane protein"
```

```
                /protein_id="CAD02374.1"
                /db_xref="GI:16503248"
                /db_xref="GOA:Q8Z5P2"
                /db_xref="InterPro:IPR002528"
                /db_xref="UniProt/TrEMBL:Q8Z5P2"
                /translation="MWFQLSSLRSILNVSTALRQVVVRTPWYAKRKSYKVLFWREITP
                LAIPIFLENTCVLLMGVLSTFLVSWLGKEAMAGVGLADSFNMVIMAFFAAIDLGTTVV
                VAFSLGKRDRRRARAAARQSLVIMTLFAVVLAVVIHYFGSEIINIVAGEATPEVKGLA
                LTYLELTVLSYPAAAIALIGSGALRGAGNTKIPLMINGGMNILNIVISSILIYGAFSW
                QGLGFAGAGLGLTISRYIGAVAIIWVLMIGFNPALRIPLKSYLKPLNFGIIWEVMGIG
                IPASIESVLFNGGKLLTQMFVAGMGTNVIAGNFIAFSVAALINLPGNALGSASTIITG
                KRLGTGQIGQAERQLRHVFWMSTIVLTAIAWGTAPFAGLFASFYTQEQDVKEVVKVLL
                WLNAAFMPIWAAAWVLPSGFKGARDVRFAMWVSMLGMWGCRVVAGYTLGIVLGMGVVG
                VWLGMFLDWAVRGALFYWRLISGRWLWRYLRVKRE"
misc_feature    complement(199. .1044)
                /gene="STY2216"
                /note="Pfam match to entry PF01554 UPF0013,
                Uncharacterized membrane protein family UPF0013, score
                273.80, E-value 2.2e-78"
tRNA            1726. .1801
                /product="tRNA-Asn"
                /note="tRNA Asn anticodon GTT, Cove score 87.13"
gene            complement(1806. .1976)
                /gene="STY2217"
CDS             complement(1806. .1976)
                /gene="STY2217"
                /note="Unknown function"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="CAD02375.1"
                /db_xref="GI:16503249"
                /db_xref="UniProt/TrEMBL:Q8Z5P1"
                /translation="MSVMAGASFSQRTARPPIKPPKTMDAKKPGKPYQVNRVKISSYF
                LNVSGLAGISRN"
gene            complement(2027. .2956)
                /gene="STY2218"
                /note="synonym: erfK"
CDS             complement(2027. .2956)
                /gene="STY2218"
                /note="Similar to Salmonella typhimurium protein erfK/srfK
                precursor SW:ERFK_SALTY (P40680) (309 aa) fasta scores:
                E(): 0, 99.0% id in 309 aa. Contains probable N-terminal
                signal sequence
                Fasta hit to YNHG_ECOLI (334 aa), 47% identity in 288 aa
                overlap
                Fasta hit to YCFS_ECOLI (320 aa), 48% identity in 307 aa
                overlap
                Fasta hit to YBIS_ECOLI (306 aa), 66% identity in 283 aa
                overlap
                Orthologue of E. coli erfK (ERFK_ECOLI); Fasta hit to
                ERFK_ECOLI (310 aa), 79% identity in 305 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="putative exported protein"
                /protein_id="CAD02376.1"
                /db_xref="GI:16503250"
                /db_xref="InterPro:IPR005490"
                /db_xref="UniProt/TrEMBL:Q8Z5P0"
                /translation="MRRITPFFPFFVLLVSHFSLAISYPLPPEGSRLVGQPVTIAVPQ
                NNTQPLESFAARYGQGLSNMLEANPGVDVFLPQSGSTLVVPQQLILPDTVREGIVVNV
                AEMRLYYYPAGTNTVEVLPIGIGQAGRETPRNWVTAVERKQDGPVWVPTANTRREYAK
                EGKTLPAMVPAGPDNPMGLYAIYIGRLYATHGTNANFGIGLRVSQGCIRLRNDDIKYL
                FDHVRVGTRVQIIDRPVKFSVEPDGSRWLEVHEPLSRNRAEFESDKKVPLPVTPVLRT
                FIKGDDVDTSRVNEVLERRSGMPVNISAGRSGL"
RBS             complement(2964. .2970)
                /note="possible RBS"
gene            complement(3034. .4104)
                /gene="STY2219"
                /note="synonym: cobT"
CDS             complement(3034. .4104)
                /gene="STY2219"
                /EC_number="2.4.2.21"
                /note="Similar to Salmonella typhimurium
                nicotinate-nucleotide--dimethylbenzimidazole
                phosphoribosyltransferase SW:COBT_SALTY (Q05603) (356 aa)
                fasta scores: E(): 0, 97.8% id in 356 aa
                Orthologue of E. coli cobT (COBT_ECOLI); Fasta hit to
                COBT_ECOLI (359 aa), 78% identity in 356 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="nicotinate-nucleotide--dimethylbenzimidazole
                phosphoribosyl transferase"
                /protein_id="CAD02377.1"
                /db_xref="GI:16503251"
                /db_xref="GOA:Q8Z5N9"
                /db_xref="InterPro:IPR003200"
                /db_xref="InterPro:IPR008281"
                /db_xref="UniProt/Swiss-Prot:Q8Z5N9"
                /translation="MQTLHALFRDIPAPDAEAMARAQQHIDGLLKPPGSLGRLETLAV
                QLAGMPGLNGTPQVGEKAVLVMCADHGVWDEGVAVSPKIVTAIQAANMTQGTTGVCVL
                AAQAGAKVHVIDVGIDAEPIPGVVDMRVARGCGNIAVGPAMSRSQAEALLLEVSRYTC
                DLAKRGVTLFGVGELGMANTTPAAAMVSVFTGSDAKEVVGIGANLPPSRIDNKVDVVR
                RAIAINQPNPRDGIDVLSKVGGFDLVGMTGVMLGAARCGLPVLLDGFLSYSAALAACQ
                IAPAVRPYLIPSHFSAEKGARIALAHLSMEPYLHMAMRLGEGSGAALAMPIVEAACAM
                FHNMGELAASNIVLPEGNANAT"
misc_feature    complement(3067. .4065)
                /gene="STY2219"
                /note="Pfam match to entry PF02277 DBI_PRT,
                Phosphoribosyltransferase, score 794.90, E-value 3.1e-235"
gene            complement(4131. .4874)
                /gene="STY2220"
                /note="synonym: cobS"
CDS             complement(4131. .4874)
                /gene="STY2220"
                /EC_number="2.-.-.-"
                /note="Similar to Salmonella typhimurium cobalamin
                [5'-phosphate] synthase cobS SW:COBS_SALTY (Q05602) (247
                aa) fasta scores: E(): 0, 97.6% id in 247 aa
                Orthologue of E. coli cobS (COBS_ECOLI); Fasta hit to
                COBS_ECOLI (247 aa), 82% identity in 247 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="cobalamin (5'-phosphate) synthase"
                /protein_id="CAD02378.1"
                /db_xref="GI:16503252"
                /db_xref="GOA:Q8Z5N8"
                /db_xref="InterPro:IPR003805"
                /db_xref="UniProt/Swiss-Prot:Q8Z5N8"
                /translation="MSKLFWAMLAFISRLPVPSRWSQGLDFEQYSRGIVMFPFIGLIL
                GGISGLIFILLQPWCGIPLAALFCILALALLTGGFHLDGLADTCDGIFSARRRERMLE
                IMRDSRLGTHGGLALIFVLLAKILVVSELALRGTPMLAALAAACAAGRGSAVLLMYRH
                RYAREEGLGNVFIGKVSGRQTCITLGLAVIVATVLLLGMQGLATMVVTLAAIFILGQL
                LKRTLGGQTGDTLGAAIELGELIFLLALL"
gene            complement(4871. .5416)
                /gene="STY2221"
                /note="synonym: cobU"
CDS             complement(4871. .5416)
                /gene="STY2221"
                /note="Similar to Salmonella typhimurium bifunctional
                cobalamin biosynthesis protein cobP [includes: cobinamide
                kinase; cobinamide phosphate guanylyltransferase] cobU
                SW:COBU_SALTY (Q05599) (181 aa) fasta scores: E(): 0,
                99.4% id in 181 aa
                Orthologue of E. coli cobU (COBU_ECOLI); Fasta hit to
                COBU_ECOLI (181 aa), 82% identity in 181 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="cobinamide kinase and guanylyltransferase"
                /protein_id="CAD02379.1"
                /db_xref="GI:16503253"
                /db_xref="GOA:Q8Z5N7"
                /db_xref="InterPro:IPR003203"
                /db_xref="UniProt/TrEMBL:Q8Z5N7"
                /translation="MMILVTGGARSGKSRHAEALISDAPQVLYIATSQILDDEMAARI
                QHHKDGRPAHWRTAECWRHLDTLITADLAPDDAILLECITTMVTNLLFALGGENDPEQ
```

WDYAAMERAIDDEIQILIAACQRCPAKVVLVTNEVGMGIVPENRLARHFRDIAGRVNQ
RLAAAADEVWLVVSGIGVKIK"

Query Match 41.6%; Score 537.8; DB 1; Length 250050;
Best Local Similarity 63.7%; Pred. No. 4.7e-73;
Matches 818; Conservative 0; Mismatches 467; Indels 0; Gaps 0;

```
Qy          3 GAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTATC 62
              |||| |    ||| | | |||||| || ||| | ||| |||| | ||| || ||| | ||
Db      75951 GAAACTGAGGTTGGTTAGAATTCCAAGACATCTTATTATTGCCGCTTCCTCTTGGCTTTC 75892

Qy         63 CAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTAGG 122
              || |||||||||||||| || ||||||| |||| ||  || |||| |     || || ||
Db      75891 AAAGATAATAATTGCCGGTGTTCAGTTAGTAAGTGTTAAATTTCTTTTAGAAATTCTTGG 75832

Qy        123 TGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCTGT 182
              |||||   ||| ||  | |||| | | |||||||| ||||| || |||| |||| ||
Db      75831 CGAAGAATCATACGCTGTATTTACTCTTTTAACTGGATTATTGGTCTGGTTTAGCATTGC 75772

Qy        183 TGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAACAA 242
              ||| |||| || ||||    | || |||||||||||||| || |  | |||  | |  ||
Db      75771 AGATATTGGGATTGGTAGTAGTCTACAAAATTATATATCTGAGTTGAAAGCTGATAGAAA 75712

Qy        243 AAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTT 302
              |   |||||||||||||| ||  | ||| | ||| |     ||   | |  | ||  ||||
Db      75711 ATCATATGATGCATATATCAAGGCCGCAGTTCATATTCTATTCGCATCCTTAATCATTTT 75652

Qy        303 TATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCA 362
              |   ||  ||| |   | || || |     | | ||  |   ||||||| |||| |||
Db      75651 AAGCTCTACATTATTCTTCTTATCAGATAAATTATCGTCACTATATCTTACTTCATTTAG 75592

Qy        363 TGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTC 422
              || | |||   |  ||||  |  |||     |||||||   || ||  |  | ||  |  |
Db      75591 CGATGAATTGAAAAACAACTCAGGAAGTTATTTTTTTATAGCAAGTATATTATTTATATT 75532

Qy        423 TATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGC 482
              || ||  | ||   | | |  |||||||||  | ||||| ||| || | ||||||||||
Db      75531 CATCGGCGTTGGGAGTGTGGTCTATAAAATATTATTTGCGGAACTGTTAGGGTGGAAAGC 75472

Qy        483 TAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTA 542
              |||| || | || |||||||||||| |  ||||| |  | |    |     ||  | ||
Db      75471 TAATATAATTAATGCATTATCTTATCTTTTAGGTTTTTTAGATGTAGTTGCGATCCATTA 75412

Qy        543 TAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTAT 602
              |    |              || |  ||  |  || |||   |||||  ||| || |  ||
Db      75411 TTTAATGCCAGATTCGAGTATTACCTTCGCTTTAGTAGCATTGTATGCTCCGGTAGCAAT 75352

Qy        603 GATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAA 662
              |  |  |    |||||   | | |  || ||| |  ||  | | |  |||  ||
Db      75351 ACTGCCCATTATATATATATCGTTTCGGTATATATATGTTCTTAAAGCGAAAGTAAACTT 75292

Qy        663 ATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATC 722
                |||| |  | |  ||    | || |||||||||  |  ||||| | |  |||||
Db      75291 TAACACCTATAAATTATTACTATCACGTTCATCAGGGTTTCTGATTTTTTCGTCCTTATC 75232

Qy        723 GATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATAT 782
              |||| | ||  | ||||| |||||||| || || ||||| |    || || |||| |||||
Db      75231 GATAATAGTTTTACAAACTGATTATATTGTGATGTCTCAGAAATTATCTGCTGCAGATAT 75172

Qy        783 TGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTAT 842
              | |  ||||||| |||||||||||||| ||||||||| | ||||||||||||||||  |
Db      75171 TATAAAATATACTGTAACGATGAAAATATTTGGTTTAATGTTTTTTATTTATACTGCGGT 75112

Qy        843 TTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACT 902
              || ||||||||||||||  |||||||||||||  |||| |||  ||| |||| ||| ||
Db      75111 ATTACAAGCATTATGGCCAGTATGTGCTGAATTACGAGTGAAAATGCAGTGGAGAAAGCT 75052

Qy        903 TAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAAT 962
              | | ||| ||      | |||||| |   |||    | | | | |||  |    ||||  |
Db      75051 GCATAGAATCATTTTCCTAAATATTATTGGTGGGGTATTTTTTATTGGTCTTGGTACGTT 74992

Qy        963 ATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCA 1022
              |||||||||| | || || ||  | || | |    |||| || || | |||  ||||| |
```

```
Db      74991 ATTTATTTATGTTTTAAAGGATTATATCTATAGCATAATTGCTAACGGTATAGATTATAA 74932

Qy       1023 AGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGA 1082
                ||  |    |   |||| | ||| | |   | |||||  ||||  | ||||||||||
Db      74931 TATTAGTGGGGTTGTTTTTGTTTTACTGGCTGTGTATTTTAGTATAAGAGTTTGGTGTGA 74872

Qy       1083 CACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTACC 1142
                || | ||| |||||| | |||||||||||  | ||||||| ||||||  |  |||| ||
Db      74871 TACATTTGCTATGTTACTTCAAAGTATGAACCAATTAAAAATTCTTTGGCTCATAGTTCC 74812

Qy       1143 ACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATCAG 1202
                || ||| ||||||||||  |  | ||||||||||||  |        ||||||
Db      74811 GTGTCAGGCATTAATTGGTGGTGTGACTCAATGGTATTTTGCAGAGCATTATGGAATAGT 74752

Qy       1203 TGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCACT 1262
                |||  |  |     ||  | ||| |||| ||  |  ||||||||||||||  | ||| |
Db      74751 TGGTATTTTATACGGACTAATTTTATCGTTCTCGCTAACTGTTTTTTGGGGATTGCCAGT 74692

Qy       1263 AACTTACTTAATTAAGGCAAATAAG 1287
                |||  |   ||||   || | |
Db      74691 GTATTATATGTATAAGAGTAAAAGG 74667
```

RESULT 7
AE016836
LOCUS AE016836 300928 bp DNA linear BCT 21-MAR-2003
DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 3 of 16 of the complete genome.
ACCESSION AE016836 AE014613
VERSION AE016836.1 GI:29136667
KEYWORDS .
SOURCE Salmonella enterica subsp. enterica serovar Typhi Ty2
ORGANISM Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 300928)
AUTHORS Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J., Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
TITLE Comparative Genomics of Salmonella enterica Serovar Typhi Strains Ty2 and CT18
JOURNAL J. Bacteriol. 185 (7), 2330-2337 (2003)
PUBMED 12644504
REFERENCE 2 (bases 1 to 300928)
AUTHORS Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J., Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2002) Laboratory of Genetics, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA

```
FEATURES             Location/Qualifiers
     source          1. .300928
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(69. .1271)
                     /gene="ackA"
                     /locus_tag="t0527"
     CDS             complement(69. .1271)
                     /gene="ackA"
                     /locus_tag="t0527"
                     /note="corresponds to STY2567 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="acetate kinase"
                     /protein_id="AAO68233.1"
                     /db_xref="GI:29136668"
                     /translation="MSSKLVLVLNCGSSSLKFAIIDAVNGDEYLSGLAECFHLPEARI
                     KWKMDGSKQEAALGAGAAHSEALNFIVNTILAQKPELSAQLTAIGHRIVHGGEKYTSS
                     VVIDESVIQGIKDSASFAPLHNPAHLIGIAEALKSFPQLKDKNVAVFDTAFHQTMPEE
```

```
                SYLYALPYSLYKEHGVRRYGAHGTSHFYVTQEAAKMLNKPVEELNIITCHLGNGGSVS
                AIRNGKCVDTSMGLTPLEGLVMGTRSGDIDPAIIFHLHDTLGMSVDQINKMLTKESGL
                LGLTEVTSDCRYVEDNYATKEDAKRAMDVYCHRLAKYIGSYTALMDGRLDAVVFTGGI
                GENAAMVRELSLGKLGVLGFEVDHERNLAARFGKSGFINKEGTRPAVVIPTNEELVIA
                QDASRLTA"
gene            1611. .2066
                /locus_tag="t0528"
CDS             1611. .2066
                /locus_tag="t0528"
                /note="corresponds to STY2566 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAO68234.1"
                /db_xref="GI:29136669"
                /translation="MSTPDNRSVNFFSLFRRGQHYAKTWPMEKRLAPVFVENRVIRMT
                RYAIRFMPPVAVFTLCWQIALGGQLGPAVATALFALSLPMQGLWWLGKRSLTPLPPSI
                LNWFYEVRGKLQEAGQALAPVEGKPDYQALADTLKRAFKQLDKTFLDDL"
gene            2149. .2643
                /locus_tag="t0529"
CDS             2149. .2643
                /locus_tag="t0529"
                /note="corresponds to STY2565 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAO68235.1"
                /db_xref="GI:29136670"
                /translation="MEMTNAQRLILSNQYKMMTMLDPTNAERYRRLQTIIERGYGLQM
                RELDREFGELTEETCRTIIDIMEMYHALHVSWTNLKDTQAIDERRVTFLGFDAATEAR
                YLGYVRFMVNIEGRYTHFDAGTHGFNAQTPMWEKYQRMLNVWHACPRQYHLSANEINQ
                IINA"
gene            2636. .3313
                /locus_tag="t0530"
CDS             2636. .3313
                /locus_tag="t0530"
                /note="corresponds to STY2564 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative phosphatase"
                /protein_id="AAO68236.1"
                /db_xref="GI:29136671"
                /translation="MPEEAHVQCKGFLFDLDGTLVDSLPAVERAWCSWADRFNLAHDE
                VLGFIHGKQAITSLRHFMAGKSEAEIAAEFTRLEQIEATETAGITALPGAVDLLNHLN
                KAGIPWAIVTSGSMPVARARHQVAGLPAPEVFVTAERVKRGKPEPDAYLLGAQLLGLA
                PQECVVVEDAPAGVLSGLAAGCHVIAVNAPADTPRLADVDFALDSLTQLSVAKQPNGD
                VVVLRKT"
gene            3389. .5215
                /locus_tag="t0531"
CDS             3389. .5215
                /locus_tag="t0531"
                /note="corresponds to STY2563 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative sodium/sulphate transporter"
                /protein_id="AAO68237.1"
                /db_xref="GI:29136672"
                /translation="MNGELIWVLSLLAIAVVLFATGKVRMDAVALFVIVAFVLSGTLT
                LSEAFSGFSDPNVILIAALFIIGDGLVRTGVATVVGTWLVKMAGSSEIKMLVLLMITV
                AGLGAFMSSTGVVAIFIPVVLSVSMHMQTSPSRLMMPLSFAGLISGMMTLVATPPNLV
                VNSELLREGLHGFNFFSVTPLGVVVLALGIVYMLVMRFMLKGDAPGQQAGKRRTFRDL
                IREYRLTGRARRLAIRPGSPMVGQRLDDLKLRERYGANVIGVERWRRFRRVIVNVNGV
                SEFRARDVLLIDMSAAEVDLREFCAEQLLEPMVLRGEYFSDQALDVGMAEISLIPKSE
                LIGKSVREIAFRTRYGLNVVGLKRDGVALEGSLADEPLLMGDIILVVGNWKLISLLGQ
                KGRDFVVLNMPVEVSEASPAHSQAPHAIFCLVLMVALMLTDEIPNPSAAIIACLLMGK
                FRCIDAESAYKAIHWPSIILIVGMMPFALALQKTGGVSLVVQGLMDIGGGYGPYMMLG
                CLFVLCAVIGLFISNTATAVLMAPIALAAAKSMGVSPYPFAMAVAMAASAAFMTPVSS
                PVNTLVLGPGNYSFSDFVKLGVPFTLIVMAVCIVMIPMLFPF"
gene            complement(5307. .5906)
                /locus_tag="t0532"
CDS             complement(5307. .5906)
                /locus_tag="t0532"
                /note="corresponds to STY2562 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative oxetanocin A biosynthetic enzyme"
                /protein_id="AAO68238.1"
                /db_xref="GI:29136673"
                /translation="MKQSHFFAHLSRMKLINRWPLMRNVRTENVSEHSLQVAMVAHAL
                AAIKNRKFGGQLNAERIALLAMYHDASEVLTGDLPTPVKYFNSQIAQEYKAIEKIAQQ
                KLVDMAPDELHDIFAPLIDENAWSKEEQAIVKQADALCAYLKCLEELSAGNNEFRLAK
                TRLEKTLELRRSQEMDYFMAVFVPSFHLSLDEISQDSPL"
gene            complement(6000. .7214)
                /locus_tag="t0533"
CDS             complement(6000. .7214)
                /locus_tag="t0533"
                /note="corresponds to STY2561 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative aminotransferase"
                /protein_id="AAO68239.1"
                /db_xref="GI:29136674"
                /translation="MSPIEKSSKLENVCYDIRGPVLKEAKRLEEEGNKVLKLNIGNPA
                PFGFEAPDEILVDVIRNLPTAQGYCDSKGLYSARKAIMQHYQARGMRDVTVEDIYIGN
                GVSELIVQAMQALLNSGDEMLVPAPDYPLWTAAVSLSSGKAVHYLCDESSDWFPDLDD
                IRAKITPRTRGIVIINPNNPTGAVYSKELLMEIVNIAREHNLIIFADEIYDKILYDDA
                EHHSIAALAPDLLTITFNGLSKTYRVAGFRQGWMVLNGPKKHAKGYIEGLEMLASMRL
                CANVPAQHAIQTALGGYQSISEFILPGGRLYEQRNRAWELINDIPGVSCVKPRGALYM
                FPRIDAKRFNIHDDQKMVLDFLLQEKVLLVQGTAFNWPWPDHFRIVTLPREDDLEMAI
                NRFGRFLSGYHQ"
gene            8146. .9084
                /gene="lrhA"
                /locus_tag="t0534"
CDS             8146. .9084
                /gene="lrhA"
                /locus_tag="t0534"
                /note="corresponds to STY2560 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="NADH dehydrogenase operon transcriptional
                regulator"
                /protein_id="AAO68240.1"
                /db_xref="GI:29136675"
                /translation="MINANRPIINLDLDLLRTFVAVADLNTFAAAAAAVCRTQSAVSQ
                QMQRLEQLVGKELFARHGRNKLLSEHGIQLLGYARKILRFNDEACSSLMFSNLQGVLT
                IGASDESADTILPFLLSRISSVYPKLALDVRVKRNAYMVDMVKSQEVDLVVTTNQPHS
                LDCLNLRTSPTHWYCAAEYVLQRGEPVPLVLLDDPSPFRDMVLETLNAAGIPWRLAYV
                ASTLLAVRAAVKAGLGVTARPVEMMSPDLRVLGAADGLPPLLDTEYMLCRDPNSQNEL
                AMVIFHAMESYQNPWHYNQFSAEGGDDPLMVEGGFE"
gene            9164. .9316
                /locus_tag="t0535"
CDS             9164. .9316
                /locus_tag="t0535"
                /note="corresponds to STY2559 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAO68241.1"
                /db_xref="GI:29136676"
                /translation="MPICTRLTAQIVKVAVNQPVAIRSSMWAKNNQNIHFMNLHVKKL
                EKLLLF"
gene            9709. .10152
                /gene="nuoA"
                /locus_tag="t0536"
CDS             9709. .10152
                /gene="nuoA"
                /locus_tag="t0536"
                /note="corresponds to STY2558 from Accession AL513382:
```

```
                    Salmonella typhi CT18"
                    /codon_start=1
                    /transl_table=11
                    /product="NADH dehydrogenase I chain A"
                    /protein_id="AAO68242.1"
                    /db_xref="GI:29136677"

  Query Match             41.6%;  Score 537.8;  DB 1;  Length 300928;
  Best Local Similarity   63.7%;  Pred. No. 4.3e-73;
  Matches  818;  Conservative    0;  Mismatches  467;  Indels    0;  Gaps    0;

Qy          3 GAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTATC 62
              |||| |    ||| | | |||||| || ||| | ||| |||| | ||| || ||| | ||
Db     276497 GAAACTGAGGTTGGTTAGAATTCCAAGACATCTTATTATTGCCGCTTCCTCTTGGCTTTC 276556

Qy         63 CAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTAGG 122
              || |||||||||||||| || |||||| |||| || || |||| |     || || ||
Db     276557 AAAGATAATAATTGCCGGTGTTCAGTTAGTAAGTGTTAAATTTCTTTTAGAAATTCTTGG 276616

Qy        123 TGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCTGT 182
              |||||   ||| ||  | |||| | | |||||||| ||||| || |||| ||||  ||
Db     276617 CGAAGAATCATACGCTGTATTTACTCTTTTAACTGGATTATTGGTCTGGTTTAGCATTGC 276676

Qy        183 TGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAACAA 242
              ||| |||| || ||||   | || |||||||||||||| || |  | |||  | |  ||
Db     276677 AGATATTGGGATTGGTAGTAGTCTACAAAATTATATATCTGAGTTGAAAGCTGATAGAAA 276736

Qy        243 AAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTT 302
              |  ||||||||||||| ||  | ||| | ||| |    ||   | |  | ||  ||||
Db     276737 ATCATATGATGCATATATCAAGGCCGCAGTTCATATTCTATTCGCATCCTTAATCATTTT 276796

Qy        303 TATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCA 362
              |   ||  ||| |   | || || |    | | ||  |   |||||| |||| |||
Db     276797 AAGCTCTACATTATTCTTCTTATCAGATAAATTATCGTCACTATATCTTACTTCATTTAG 276856

Qy        363 TGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTC 422
              || | |||  |  |||| |  |||     |||||||   || || |  | || |  |
Db     276857 CGATGAATTGAAAAACAACTCAGGAAGTTATTTTTTTATAGCAAGTATATTATTTATATT 276916

Qy        423 TATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGC 482
              || ||  | ||   | | |  ||||||||| | ||||| ||| || | ||||||||||
Db     276917 CATCGGCGTTGGGAGTGTGGTCTATAAAATATTATTTGCGGAACTGTTAGGGTGGAAAGC 276976

Qy        483 TAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTA 542
              |||| || | || |||||||||||||  |||||  |  | |   |  |    ||  | ||
Db     276977 TAATATAATTAATGCATTATCTTATCTTTTAGGTTTTTTAGATGTAGTTGCGATCCATTA 277036

Qy        543 TAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTAT 602
              |    |          || |  ||  |  || |||   |||||  ||| || |  ||
Db     277037 TTTAATGCCAGATTCGAGTATTACCTTCGCTTTAGTAGCATTGTATGCTCCGGTAGCAAT 277096

Qy        603 GATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAA 662
              |  |  |    |||||    | | | || ||| |  ||  | | |  |||  ||
Db     277097 ACTGCCCATTATATATATATCGTTTCGGTATATATATGTTCTTAAAGCGAAAGTAAACTT 277156

Qy        663 ATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATC 722
                  ||||  |  | |  ||     | || |||||||| |  |||||  | |  |||||
Db     277157 TAACACCTATAAATTATTACTATCACGTTCATCAGGGTTTCTGATTTTTTCGTCCTTATC 277216

Qy        723 GATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATAT 782
              |||| | ||  | ||||| |||||||| || || |||||  |    || || |||| |||||
Db     277217 GATAATAGTTTTACAAACTGATTATATTGTGATGTCTCAGAAATTATCTGCTGCAGATAT 277276

Qy        783 TGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTAT 842
              | |  ||||||| ||||||||||||||| ||||||||| | ||||||||||||||||  |
Db     277277 TATAAAATATACTGTAACGATGAAAATATTTGGTTTAATGTTTTTTATTTATACTGCGGT 277336

Qy        843 TTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACT 902
              || ||||||||||||||  ||||||||||||||  |||| |||  ||| |||| ||| ||
Db     277337 ATTACAAGCATTATGGCCAGTATGTGCTGAATTACGAGTGAAAATGCAGTGGAGAAAGCT 277396

Qy        903 TAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAAT 962
                | | ||| ||     | |||||| |   |||   | | | | ||| |    ||||  |

Db     277397 GCATAGAATCATTTTCCTAAATATTATTGGTGGGGTATTTTTTATTGGTCTTGGTACGTT 277456

Qy        963 ATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCA 1022
              |||||||||| | || || ||  | || | |    |||| || || | |||  ||||| |
Db     277457 ATTTATTTATGTTTTAAAGGATTATATCTATAGCATAATTGCTAACGGTATAGATTATAA 277516

Qy       1023 AGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGA 1082
              ||  |    |  |||| | ||| | |   | |||||  ||||  | ||||||||||
Db     277517 TATTAGTGGGGTTGTTTTTGTTTTACTGGCTGTGTATTTTAGTATAAGAGTTTGGTGTGA 277576

Qy       1083 CACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTACC 1142
              || | ||| |||||| | ||||||||||  | |||||||| |||||| |  |||| ||
Db     277577 TACATTTGCTATGTTACTTCAAAGTATGAACCAATTAAAAATTCTTTGGCTCATAGTTCC 277636

Qy       1143 ACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATCAG 1202
                || ||| ||||||||||  |  | ||||||||||||| |           ||||||
Db     277637 GTGTCAGGCATTAATTGGTGGTGTGACTCAATGGTATTTTGCAGAGCATTATGGAATAGT 277696

Qy       1203 TGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTTGGGGGCTTCCACT 1262
              |||  |  |    ||  | ||| |||| ||  |  ||||||||||||||  | ||| |
Db     277697 TGGTATTTTATACGGACTAATTTTATCGTTCTCGCTAACTGTTTTTTTGGGGATTGCCAGT 277756

Qy       1263 AACTTACTTAATTAAGGCAAATAAG 1287
                |||  |   ||||   || | |
Db     277757 GTATTATATGTATAAGAGTAAAAGG 277781


RESULT 8
CP000026_08
WPCOMMENT
  Sequence split into 46 fragments  LOCUS CP000026 Accession CP000026
     Fragment Name      Begin     End
     CP000026_00            1    110000
     CP000026_01       100001    210000
     CP000026_02       200001    310000
     CP000026_03       300001    410000
     CP000026_04       400001    510000
     CP000026_05       500001    610000
     CP000026_06       600001    710000
     CP000026_07       700001    810000
     CP000026_08       800001    910000
     CP000026_09       900001   1010000
     CP000026_10      1000001   1110000
     CP000026_11      1100001   1210000
     CP000026_12      1200001   1310000
     CP000026_13      1300001   1410000
     CP000026_14      1400001   1510000
     CP000026_15      1500001   1610000
     CP000026_16      1600001   1710000
     CP000026_17      1700001   1810000
     CP000026_18      1800001   1910000
     CP000026_19      1900001   2010000
     CP000026_20      2000001   2110000
     CP000026_21      2100001   2210000
     CP000026_22      2200001   2310000
     CP000026_23      2300001   2410000
     CP000026_24      2400001   2510000
     CP000026_25      2500001   2610000
     CP000026_26      2600001   2710000
     CP000026_27      2700001   2810000
     CP000026_28      2800001   2910000
     CP000026_29      2900001   3010000
     CP000026_30      3000001   3110000
     CP000026_31      3100001   3210000
     CP000026_32      3200001   3310000
     CP000026_33      3300001   3410000
     CP000026_34      3400001   3510000
     CP000026_35      3500001   3610000
     CP000026_36      3600001   3710000
     CP000026_37      3700001   3810000
     CP000026_38      3800001   3910000
     CP000026_39      3900001   4010000
     CP000026_40      4000001   4110000
```

```
     CP000026_41       4100001   4210000
     CP000026_42       4200001   4310000
     CP000026_43       4300001   4410000
     CP000026_44       4400001   4510000
     CP000026_45       4500001   4585229
Continuation (9 of 46) of CP000026 from base 800001 (CP000026 Salmonella enterica subsp.

  Query Match           41.5%;  Score 536.2;  DB 1;  Length 110000;
  Best Local Similarity  63.6%;  Pred. No. 1.1e-72;
  Matches  817;  Conservative    0;  Mismatches  468;  Indels    0;  Gaps    0;

Qy          3 GAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTATC 62
              |||| |    ||| | | |||||| || ||| | ||| |||| | ||| || ||| | ||
Db      69681 GAAACTGAGGTTGGTTAGAATTCCAAGACATCTTATTATTGCCGCTTCCTCTTGGCTTTC 69740

Qy         63 CAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTAGG 122
              || |||||||||||||| || |||||| |||| ||  || |||| |    || || ||
Db      69741 AAAGATAATAATTGCCGGTGTTCAGTTAGTAAGTGTTAAATTTCTTTTGGAAATTCTTGG 69800

Qy        123 TGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCTGT 182
              |||||   ||| ||  | |||| | | |||||||| ||||| || |||| ||||  ||
Db      69801 CGAAGAATCATACGCTGTATTTACTCTTTTAACTGGATTATTGGTCTGGTTTAGCATTGC 69860

Qy        183 TGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAACAA 242
              ||| |||| || ||||   | || |||||||||||||| || |  | |||  | |  ||
Db      69861 AGATATTGGGATTGGTAGTAGTCTACAAAATTATATATCTGAGTTGAAAGCTGATAGAAA 69920

Qy        243 AAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTT 302
              |  |||||||||||| ||  | ||| | ||| |    ||  | |  | ||  ||||
Db      69921 ATCATATGATGCATATATCAAGGCCGCAGTTCATATTCTATTCGCATCCTTAATCATTTT 69980

Qy        303 TATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCA 362
              |   ||  ||| |   | || || |    | | ||  |   ||||||| |||| |||
Db      69981 AAGCTCTACATTATTCTTCTTATCAGATAAATTATCGTCACTATATCTTACTTCATTTAG 70040

Qy        363 TGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTC 422
              || | |||  |  ||||  |  |||      |||||||   || || |  | || |  |
Db      70041 CGATGAATTGAAAAACAACTCAGGAAGTTATTTTTTTATAGCAAGTATATTATTTATATT 70100

Qy        423 TATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGC 482
              || ||  | ||   | | |  |||||||| | ||||| ||| || | ||||||||||
Db      70101 CATCGGCGTTGGGAGTGTGGTCTATAAAATATTATTTGCGGAACTGTTAGGGTGGAAAGC 70160

Qy        483 TAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTA 542
              |||| || | || |||||||||||| |  ||||| |  | |   |  |    ||  | ||
Db      70161 TAATATAATTAATGCATTATCTTATCTTTTAGGTTTTTTAGATGTAGTTGCGATCCATTA 70220

Qy        543 TAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTAT 602
              |    |           || |  ||  |  || |||   |||||  ||| || |  ||
Db      70221 TTTAATGCCAGATTCGAGTATTACCTTCGCTTTAGTAGCATTGTATGCTCCGGTAGCAAT 70280

Qy        603 GATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAA 662
              |  |  |    |||||     | | | || ||| |   ||  | | |  |||  ||
Db      70281 ACTGCCCATTATATATATATCGTTTCGGTATATATATGTTCTTAAAGCGAAAGTAAACTT 70340

Qy        663 ATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATC 722
                 |||| |  |  |  ||     | || |||||||||  |  |||||  |  |  |||||
Db      70341 TAACACCTATAAATTATTACTATCACGTTCATCAGGGTTTCTGATTTTTTCGTCCTTATC 70400

Qy        723 GATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATAT 782
              |||| | ||  | ||||| |||||||| || || ||||| |   || || |||| |||||
Db      70401 GATAATAGTTTTACAAACTGATTATATTGTGATGTCTCAGAAATTATCTGCTGCAGATAT 70460

Qy        783 TGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTAT 842
              | |  |||||||  |||||||||||||| |||||||| | ||||||||||||||||  |
Db      70461 TATAAAATATACTGTAACGATGAAAATATTTGGTTTAATGTTTTTTATTTATACTGCGGT 70520

Qy        843 TTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACT 902
              || |||||||||||||  |||||||||||||  |||| |||  ||| |||| ||| ||
Db      70521 ATTACAAGCATTATGGCCAGTATGTGCTGAATTACGAGTGAAAATGCAGTGGAGAAAGCT 70580

Qy        903 TAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAAT 962
              | | ||| ||     | ||||||  |   |||   | | | | ||| |    ||||  |

Db      70581 GCATAGAATCATTTTCCTAAATATTATTGGTGGGGTATTTTTTATTGGTCTTGGTACGTT 70640

Qy        963 ATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCA 1022
              |||||||||| | || || |   | || | |    |||| || || | |||  ||||| |
Db      70641 ATTTATTTATGTTTTAAAGGGTTATATCTATAGCATAATTGCTAACGGTATAGATTATAA 70700

Qy       1023 AGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGA 1082
                || |    |   |||| | ||| | |   | |||||  ||||  | ||||||||||
Db      70701 TATTAGTGAGGTTGTTTTTGTTTTACTGGCTGTGTATTTTAGTATAAGAGTTTGGTGTGA 70760

Qy       1083 CACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTACC 1142
              || | ||| |||||| | ||||||||||  | |||||||| |||||| |  |||| ||
Db      70761 TACATTTGCTATGTTACTTCAAAGTATGAACCAATTAAAAATTCTTTGGCTCATAGTTCC 70820

Qy       1143 ACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATCAG 1202
                || ||| ||||||||||  |  | |||||||||||| |          ||||||
Db      70821 GTGTCAGGCATTAATTGGTGGTGTGACTCAATGGTATTTTGCAGAGCATTATGGAATAGT 70880

Qy       1203 TGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCACT 1262
              |||  |  |    ||  | ||| |||| ||  |  ||||||||||||||  | ||| |
Db      70881 TGGTATTTTATACGGACTAATTTTATCGTTCTCGCTAACTGTTTTTTGGGGATTGCCAGT 70940

Qy       1263 AACTTACTTAATTAAGGCAAATAAG 1287
                 |||  |   ||||   || | |
Db      70941 GTATTATATGTATAAGAGTAAAAGG 70965
```

```
RESULT 9
AF461769
LOCUS       AF461769               19457 bp    DNA     linear   BCT 26-SEP-2002
DEFINITION  Yersinia pseudotuberculosis serogroup IVB O-antigen gene cluster,
            complete sequence.
ACCESSION   AF461769 U29691
VERSION     AF461769.1  GI:23321111
KEYWORDS    .
SOURCE      Yersinia pseudotuberculosis
  ORGANISM  Yersinia pseudotuberculosis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Yersinia.
REFERENCE   1  (bases 1 to 19457)
  AUTHORS   Hobbs,M. and Reeves,P.R.
  TITLE     Genetic organisation and evolution of Yersinia pseudotuberculosis
            3,6-dideoxyhexose biosynthetic genes
  JOURNAL   Biochim. Biophys. Acta 1245 (3), 273-277 (1995)
   PUBMED   8541300
REFERENCE   2  (bases 1 to 19457)
  AUTHORS   Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C.,
            Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
  TITLE     Bacterial polysaccharide synthesis and gene nomenclature
  JOURNAL   Trends Microbiol. 4 (12), 495-503 (1996)
   PUBMED   9004408
REFERENCE   3  (bases 1 to 19457)
  AUTHORS   Pacinelli,E., Wang,L. and Reeves,P.R.
  TITLE     Relationship of Yersinia pseudotuberculosis O antigens IA, IIA, and
            IVB: the IIA gene cluster was derived from that of IVB
  JOURNAL   Infect. Immun. 70 (6), 3271-3276 (2002)
   PUBMED   12011023
REFERENCE   4  (bases 1 to 19457)
  AUTHORS   Hobbs,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-1995) Department of Microbiology, University of
            Sydney, Building G08, Sydney, NSW 2006, Australia
REFERENCE   5  (bases 1 to 19457)
  AUTHORS   Pacinelli,E., Wang,L. and Reeves,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) Microbiology, University of Sydney, Sydney,
            NSW 2006, Australia
  REMARK    Nucleotide and amino acid sequences updated by submitter
COMMENT     On Sep 26, 2002 this sequence version replaced gi:1223958.
FEATURES             Location/Qualifiers
     source          1. .19457
                     /organism="Yersinia pseudotuberculosis"
                     /mol_type="genomic DNA"
```

```
                /isolate="M452"
                /db_xref="taxon:633"
                /note="serogroup: IVB"
     gene       243. .1232
                /gene="ddhD"
     CDS        243. .1232
                /gene="ddhD"
                /note="DdhD"
                /codon_start=1
                /transl_table=11
                /product="CDP-6-deoxy-delta-3,4-glucoseen reductase"
                /protein_id="AAN23052.1"
                /db_xref="GI:23321112"
                /translation="MSLNVKLHPSGIIFTSDGTSTILDAALDSNIHIEYSCKDGTCGS
                CKAILISGEVDSAENTFLTEEDVAKGAILTCCSKAKSDIELDVNYYPELSHIQKKTYP
                CKLDSIEFVGEDIAILSLRLPPTAKIQYLAGQYIDLIINGQRRSYSIANAPGGNGNIE
                LHVRKVVNGVFSNIIFNELKLQQLLRIEGPQGTFFVREDNLPIVFLAGGTGFAPVKSM
                VEALINKNDQRQVHIYWGMPAGHNFYSDIANEWAIKHPNIHYVPVVSGDDSTWTGATG
                FVHQAVLEDIPDLSLFNVYACGSLAMITAARNDFINHGLAENKFFSDAFVPSK"
     gene       1258. .2043
                /gene="ddhA"
     CDS        1258. .2043
                /gene="ddhA"
                /note="DdhA"
                /codon_start=1
                /transl_table=11
                /product="glucose-1-P-cytidylyltransferase"
                /protein_id="AAN23053.1"
                /db_xref="GI:23321113"
                /translation="MEIQVKAVILAGGLGTRLSEETVVKPKPMVEIGGKPILWHIMKL
                YSSYGINDFVICCGYKGYVIKEYFANYFMHMSDITFCMRDNEMIVHQKRVEPWNVTLV
                DTGEDSMTGGRLRRVKDYVKDDEAFCFTYGDGVSDVNIAELIEFHKSHGKQATLTATY
                PPGRFGALDIKDKQVRSFKEKPKGDGALINGGYFVLSPKVIDLIDGDKSTWEQEPLMT
                LAAQGELMAFEHAGFWQPMDTLRDKIYLHELWEEGRAPWKVWE"
     gene       1946. .3121
                /gene="ddhB"
     CDS        1946. .3121
                /gene="ddhB"
                /note="DdhB"
                /codon_start=1
                /transl_table=11
                /product="CDP-glucose-4,6-dehydratase"
                /protein_id="AAN23054.1"
                /db_xref="GI:23321114"
                /translation="MPVSGSQWIRCVTRSICMNYGKKAGHLGRYGNNKMINNSFWQGK
                RVFVTGHTGFKGGWLSLWLQTMGATVKGYSLAPPTVPSLFETARVADGMQSEIGDIRD
                QNKLLESIREFQPEIVFHMAAQPLVRLSYSEPVETYSTNVMGTVYLLEAIRHVGGVKA
                VVNITSDKCYDNKEWIWGYRENEAMGGYDPYSNSKGCAELVTSSYRNSFFNPANYGQH
                GTAVATVRAGNVIGGGDWALDRIVPDILRAFEQSQPVIIRNPHAIRPWQHVLEPLSGY
                LLLAQKLYTDGAEYAEGWNFGPNDADATPVKNIVEQMVKYWGEGASWQLDGNAHPHEA
                HYLKLDCSKAKMQLGWHPRWNLNTTLEYIVGWHKNWLSGTDMHEYSITEINNYMNTK"
     gene       3139. .4452
                /gene="ddhC"
     CDS        3139. .4452
                /gene="ddhC"
                /note="DdhC; formerly known as rfbH"
                /codon_start=1
                /transl_table=11
                /product="CDP-4-keto-6-deoxy-D-glucose-3-dehydrase"
                /protein_id="AAN23066.1"
                /db_xref="GI:23321126"
                /translation="MSQEELRQQIAELVAQYAETAMAPKPFEAGKSVVPPSGKVIGTK
                ELQLMVEASLDGWLTTGRFNDAFEKKLGEYLGVPYVLTTTSGSSANLLALTALTSPKL
                GVRALKPGDEVITVAAGFPTTVNPTIQNGLIPVFVDVDIPTYNVNASLIEAAVSDKTK
                AIMIAHTLGNLFDLAEVRRVADKYNLWLIEDCCDALGSTYDGKMAGTFGDIGTVSFYP
                AHHITMGEGGAVFTQSAELKSIIESFRDWGRDCYCAPGCDNTCKKRFGQQLGSLPFGY
                DHKYTYSHLGYNLKITDMQAACGLAQLERIEEFVEKRKANFKYLKDALQSCADFIELP
                EATENSDPSWFGFPITLKEDSGVSRIDLVKFLDEAKVGTRLLFAGNLTRQPYFHDVKY
                RVVGELTNTDRIMNQTFWIGIYPGLTHDHLDYVVSKFEEFFGLNF"
     repeat_region 4463. .4478
                /rpt_type=inverted
     gene       4488. .5345
                /gene="prt"
     CDS        4488. .5345
                /gene="prt"
                /note="Prt; formerly known as rfbS"
                /codon_start=1
                /transl_table=11
                /product="CDP-paratose synthetase"
                /protein_id="AAN23067.1"
                /db_xref="GI:23321127"
                /translation="MKILITGVSGYLGSQLANALMLEHEVAGTVRAGSVCNRITDIGN
                VNLINVTDSGWIDKVLSFSPDVVINTVALYGRKGELLSELVDANIQFPLRILEMLVST
                GKGLFFQCGTSLPADVSQYALTKNQFTELAREYCNKFSGKFIELKLEHFFGPFDDSTK
                FTTYVINSCRSHSDLKLTAGLQRRDFIYINDLINAFKIMISKSESLISGESISIGSGH
                AVTIKEFVETVAKMTSYQGNLQFGAIPTRENELMYSCASLARIQELGWLCQYSLNSAI
                KDILHKEYK"
     gene       5342. .6358
                /gene="tyv"
     CDS        5342. .6358
                /gene="tyv"
                /note="Tyv; formerly known as rfbE"
                /codon_start=1
                /transl_table=11
                /product="CDP-tyvelose epimerase"
                /protein_id="AAN23068.1"
                /db_xref="GI:23321128"
                /translation="MKLLITGGCGFLGSNLASHAIKSGMEVIVFDNLSRYGSSDNLKW
                LQSIGGFTYVHGDIRNRNDITRLIQKFKPDNIFHLAGQVAMTTSIDNPQMDFEVNVGG
                TFNLLEAIRLFNPECGIIYSSTNKVYGDLEQFTYRETDTRYECIEMPDGFDESTQLTF
                HSPYGCSKGAADQYMLDYARIYGLKTVVFRHSSMYGGRQFSTYDQGWVGWFCQKAIEA
                SRGVNSPFTISGNGKQVRDVLHAEDIISLYFSTLSNLERVKGNAFNIGGTIEHSLSLL
                ELFSLLEKYTETELKYTRIPVRESDQKVFVANINKISESTGWIPKVSSESGIKIMLDW
                VETV"
     gene       6390. .7703
                /gene="wzx"
     CDS        6390. .7703
                /gene="wzx"
                /note="Wzx; formerly known as rfbX"
                /codon_start=1
                /transl_table=11
                /product="O-unit flippase-like protein"
                /protein_id="AAN23069.1"
                /db_xref="GI:23321129"
                /translation="MRVPTHIIVAASAWGSRLVSIFIQFYSIKILLDLLGTEGYAVFT
                VIGSLVGWFLLADPGLGNSLQNQISYRRANHQEYQDLVLSAVIAIIPIFILFIILILT
                LSPYISEFLLGGFDFLNNNQRSNIFKAASFIFLTTSIGNLAYKIWFSEHKGWISNIIP
                ALSSIVGLIFLMRLPSDGSNISEDIIFSIYCFYIPAAFFGVTSTLFKVIPYLKCKNFL
                NKLTLYALIKNGGGFFLFSLLSAFVLQVDYIVMSQTLVERDLVTYNIMSKTFGLINFI
                YAALLQSLWPVCAEASSKLRFDNFYKIEKKYISFGFIIVIVSSFVIFLLKDFIVNILA
                PGKDFYFPISLIILFSFYQVVRVWTDTYAMFLMSIGKLKPLWISVPFQAVLSGSLQWV
                GAVNYGLVGLLCGLIASFLITVSWWLPFSFRSTVDRIVKDKRLDE"
     gene       7696. .8733
                /gene="wbyA"
     CDS        7696. .8733
                /gene="wbyA"
                /note="WbyA"
                /codon_start=1
                /transl_table=11
                /product="putative tyvelose transferase"
                /protein_id="AAN23055.1"
                /db_xref="GI:23321115"
                /translation="MSNIKISFCIPTYNRSELLSELIESIVSQCKNRNDIEICISDNA
                SSDNTSEMIAKWKESTDTPIVYKVNKENIGPDRNYLAAVDLASGEYCWLFGSDDRLLS
                DAIIKMDEHINEGHDIYIVDRIECDFNMKPFSHRTWMSTGSKLYNTSKRTDLIEYFNN
                SYAVGAVFSYLSSIIVKRSRWNRVKFNPKFIGSCYAHVFYLLDIVNNGAELKYVKSAF
                VLSRGGNDSFAANGLIKRIEIDFDGYFLLGNEPFSNDFYLNSKFKAILLKEHPVIHTC
                IKTAMIGEFQDLNKIKRFYIELGIAPYYVSIIYFFKPIYQLANKMKLSFFYKKLRKKI
                TFKKGSFNSID"
     gene       8768. .9877
                /gene="wbyB"
     CDS        8768. .9877
                /gene="wbyB"
                /note="WbyB"
                /codon_start=1
                /transl_table=11
```

/product="putative 6-deoxy-D-mannoheptose transferase"
/protein_id="AAN23056.1"
/db_xref="GI:23321116"
/translation="MNILIVYLGRKGGGAHYTYELVNELSKKTNVSLLVSKQAENIKS
YESLNISLNTINTFNGMFSFILSSFKLPFIKNKIKKIILDEKIDIVISTMTHVWNGFI
FTPNFMMKIPYILTIHDAVPHSGEENFILNLLNNKDIKNANAYIALTEHVKNQFINIY
GDKKDICILPLGSFNFKSGEIKSWCGVRPLRILFFGRIHEYKGIDLLLDSMIILQEKY
DLELSIYGQGDLSDYVDKIDVIKKINIENRWIGDSEVGDILQAHDVCVIPYKDASQSG
VIPVAQSAGLPTIVNPVDGLIEQIEKDKTSLVTDRVDSVSLALSIEKIIKNPELVNKL
SIGSLEYASRKLSWVPIVNKLITYSKETLNSSDRS"

Query Match 19.8%; Score 256.6; DB 1; Length 19457;
Best Local Similarity 51.3%; Pred. No. 5.2e-30;
Matches 654; Conservative 0; Mismatches 609; Indels 12; Gaps 2;

```
Qy        23 TTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTATCCAAAATAATAATTGCCGGGG 82
             | || |  ||| |||| |||||||      |||||          || ||  |
Db      6397 TACCTACACATATAATAGTTGCAGCAAGTGCATGGGGGAGTCGGTTAGTATCTATATTTA 6456

Qy        83 TGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCT 142
             | || ||    ||||||  |  | | |  |  | | ||| |  ||   ||||||  | |
Db      6457 TCCAATTTTATAGTATTAAAATTTTATTGGATCTATTGGGTACGGAAGGATATGCTGTTT 6516

Qy       143 TTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAG 202
             |||   |  |   | |  || | |  ||||       ||  ||||||||  | ||||
Db      6517 TTACCGTTATTGGTAGCCTAGTTGGTTGGTTCCTGCTTGCGGATTTTGGTCTCGGTAATA 6576

Qy       203 GACTGCAAAATTATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTA 262
             | || || ||| | |||||| |  | ||||| ||  |  ||   ||| | |  |   ||
Db      6577 GCCTTCAGAATCAAATATCATATAGGAGAGCAAATCATCAAGAATATCAAGATTTAGTTT 6636

Qy       263 AATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTT 322
              ||| ||| ||      |   |      | ||||||| || ||   |||| | |  |  |
Db      6637 TATCTGCAGTAATAGCTATAATACCTATTTTTATTTTATTCATAATTTTAATATTAACAT 6696

Qy       323 TTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAA 382
             | ||      || || |      |||  |  ||| ||   ||   |  |  | |
Db      6697 TGTCACCTTATATATCTGAGTTCCTCCTTGGTGGATTTGATTTTTTAAATAACAATCAGA 6756

Qy       383 CCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTG 442
              || |  | |||    | ||| | |  |||||   |||     || ||   | | |
Db      6757 GAAGCAATATATTTAAAGCGGCATCTTTTATTTTCTTAACTACATCCATTGGTAATTTAG 6816

Qy       443 CTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTAT 502
             | |||||||||   ||| | ||       || ||||   |||||  || |   |||||| |
Db      6817 CATATAAAATATGGTTTTCAGAGCATAAAGGATGGATTTCTAATATAATTCCCGCATTGT 6876

Qy       503 CTTATATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTT---- 558
             | | |||  | ||  |  |    ||| |     ||   | |   || | |    ||
Db      6877 CATCTATTGTTGGACTCATTTTTTTGATGAGACTACCATCTGATGGAAGCAATATTAGTG 6936

Qy       559 --GACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCT 616
               || || |  || ||  |        | ||| ||||||  |   |  ||    |    |
Db      6937 AAGATATTATTTTTTCTATTTATTGTTTTTATATTCCAGCTGCATTTTTTGGTGTAACAT 6996

Qy       617 ATATTGTATATAGATACATAAAGCTTTATCATGTTAAAA------CAACAAAATCTCATT 670
              || | ||| || |  ||||     ||   |  |||||       ||  || |
Db      6997 CTACTTTATTTAAAGTCATACCTTATTTAAAATGTAAAAATTTTTTAAATAAGTTGACGC 7056

Qy       671 ATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGG 730
              |   ||  |  ||   |        || ||||| || |||| | || ||||   | | |
Db      7057 TTTATGCTCTAATAAAAAATGGAGGCGGTTTTTTCCTATTTAGTCTACTATCAGCATTTG 7116

Qy       731 TGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAAT 790
             | ||||||   || || || || || || || | | ||          ||| | ||    |
Db      7117 TACTTCAAGTTGACTACATTGTTATGTCACAGACGTTAGTAGAAAGAGATTTGGTAACGT 7176

Qy       791 ATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAG 850
             |||   ||| ||  |||| |||||||||| |    || || ||| | ||| | ||||||
Db      7177 ATAACATAATGAGTAAAACTTTTGGTTTAATTAACTTCATATATGCGGCTTTATTGCAAT 7236

Qy       851 CATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAA 910
              || | |||||| ||||||| ||    ||    ||||     |   | ||  || | ||||
Db      7237 CACTTTGGCCTGTATGTGCAGAGGCAAGTTCAAAACTAAGGTTTGATAATTTTTATAAAA 7296

Qy        911 TGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTT 970
              |   |      |||| | | |||| |   |   |||| ||| | || || |  ||||||
Db      7297 TAGAAAAAAAATATATATCGTTTGGTTTTATTATTGTTATTGTAAGTTCATTTGTTATTT 7356

Qy        971 ATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTA 1030
              |||||| |||||    ||  |     || | ||    |  |   | |   |  |  |||
Db      7357 TTTTATTAAAAGACTTTATTGTAAATATACTCGCACCTGGCAAAGACTTCTATTTCCCTA 7416

Qy       1031 TTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATG 1090
              ||   |  |||| ||  ||  | | |||       |  | || |||  ||| || ||||
Db      7417 TTAGTTTGATTATATTGTTTTCCTTTTATCAAGTGGTAAGGGTGTGGACTGATACATATG 7476

Qy       1091 CAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTACCACTACAAG 1150
              | ||||  |   ||||  | | ||||||  ||| ||||    || ||  | ||||
Db      7477 CTATGTTTCTCATGAGTATTGGTAAATTAAAACCACTATGGATTAGTGTTCCCTTTCAAG 7536

Qy       1151 CAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATCAGTGGAGTGC 1210
              |  || |  |||   |  ||||||  |  | |    |   |||  |   ||| | | |
Db      7537 CTGTATTAAGTGGTTCACTTCAATGGGTTGGTGCAGTAAATTATGGTTTAGTAGGATTAC 7596

Qy       1211 TGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCACTAACTTACT 1270
              |   ||| || ||   ||| ||      | || ||||  ||| || | ||  | | ||
Db      7597 TTTGTGGATTAATAGCATCCTTCTTAATTACCGTTTCATGGTGGTTACCTTTTAGTTTTC 7656

Qy       1271 TAATTAAGGCAAATA 1285
              |  ||  | | |||
Db      7657 GATCTACAGTAGATA 7671
```

RESULT 10
AF461770
LOCUS AF461770 20125 bp DNA linear BCT 26-SEP-2002
DEFINITION Yersinia pseudotuberculosis serogroup IIA O-antigen gene cluster, complete sequence.
ACCESSION AF461770 L01777
VERSION AF461770.1 GI:23321130
KEYWORDS .
SOURCE Yersinia pseudotuberculosis
ORGANISM Yersinia pseudotuberculosis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.
REFERENCE 1 (bases 1 to 20125)
AUTHORS Kessler,A.
TITLE The 3,6 dideoxyhexose pathway genes of Yersinia pseudotuberculosis serogroup IIA
JOURNAL Thesis (1993) Department of Microbiology, University of Sydney, Australia
REFERENCE 2 (bases 1 to 8845)
AUTHORS Kessler,A.C., Haase,A. and Reeves,P.R.
TITLE Molecular analysis of the 3,6-dideoxyhexose pathway genes of Yersinia pseudotuberculosis serogroup IIA
JOURNAL J. Bacteriol. 175 (5), 1412-1422 (1993)
PUBMED 8444803
REFERENCE 3 (bases 1 to 20125)
AUTHORS Hobbs,M. and Reeves,P.R.
TITLE The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters
JOURNAL Mol. Microbiol. 12 (5), 855-856 (1994)
PUBMED 8052136
REFERENCE 4 (bases 1 to 20125)
AUTHORS Liu,D., Lindqvist,L. and Reeves,P.R.
TITLE Transferases of O-antigen biosynthesis in Salmonella enterica: dideoxyhexosyltransferases of groups B and C2 and acetyltransferase of group C2
JOURNAL J. Bacteriol. 177 (14), 4084-4088 (1995)
PUBMED 7541787
REFERENCE 5 (bases 1 to 20125)
AUTHORS Hobbs,M. and Reeves,P.R.
TITLE Genetic organisation and evolution of Yersinia pseudotuberculosis 3,6-dideoxyhexose biosynthetic genes
JOURNAL Biochim. Biophys. Acta 1245 (3), 273-277 (1995)

```
   PUBMED   8541300
REFERENCE   6  (bases 1 to 20125)
  AUTHORS   Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C.,
            Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
  TITLE     Bacterial polysaccharide synthesis and gene nomenclature
  JOURNAL   Trends Microbiol. 4 (12), 495-503 (1996)
   PUBMED   9004408
REFERENCE   7  (bases 1 to 20125)
  AUTHORS   Pacinelli,E., Wang,L. and Reeves,P.R.
  TITLE     Relationship of Yersinia pseudotuberculosis O antigens IA, IIA, and
            IVB: the IIA gene cluster was derived from that of IVB
  JOURNAL   Infect. Immun. 70 (6), 3271-3276 (2002)
   PUBMED   12011023
REFERENCE   8  (bases 1 to 20125)
  AUTHORS   Hobbs,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1993) Department of Microbiology, University of
            Sydney, Building G08, Sydney, NSW 2006, Australia
REFERENCE   9  (bases 1 to 20125)
  AUTHORS   Pacinelli,E., Wang,L. and Reeves,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) Microbiology, University of Sydney, Sydney,
            NSW 2006, Australia
  REMARK    Nucleotide and amino acid sequences updated by submitter
COMMENT     On Sep 26, 2002 this sequence version replaced gi:896325.
FEATURES             Location/Qualifiers
     source          1. .20125
                     /organism="Yersinia pseudotuberculosis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:633"
                     /note="serogroup: IIA"
     repeat_region   1. .10009
                     /note="remnants of an IS related to IS630"
                     /insertion_seq="IS630-like"
     gene            <1. .235
                     /gene="hemH"
     CDS             <1. .235
                     /gene="hemH"
                     /codon_start=2
                     /transl_table=11
                     /product="ferrochelatase"
                     /protein_id="AAB49397.1"
                     /db_xref="GI:155494"
                     /translation="TPYTDETLKSLPSQGVKHIQLICPGFSADCLETLEEIKEQNREF
                     FLHAGGEKFEYIPALNDDEGHIALLEQLIRHNI"
     misc_feature    586. .623
                     /note="JUMPstart"
                     /citation=[3]
     gene            833. .1822
                     /gene="ddhD"
     CDS             833. .1822
                     /gene="ddhD"
                     /note="DdhD; formerly known as rfbI"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-6-deoxy-delta-3,4-glucoseen reductase"
                     /protein_id="AAB49398.1"
                     /db_xref="GI:155495"
                     /translation="MSLNVKLHPSGIIFTSDGTSTILDAALDSNIHIEYSCKDGTCGS
                     CKAILISGEVDSAENTFLTEEDVAKGAILTCCSKAKSDIELDVNYYPELSHIQKKTYP
                     CKLDSIEFIGEDIAILSLRLPPTAKIQYLAGQYIDLIINGQRRSYSIANAPGGNGNIE
                     LHVRKVVNGVFSNIIFNELKLQQLLRIEGPQGTFFVREDNLPIVFLAGGTGFAPVKSM
                     VEALINKNDQRQVHIYWGMPAGHNFYSDIANEWAIKHPNIHYVPVVSGDDSTWTGATG
                     FVHQAVLEDIPDLSLFNVYACGSLAMITAARNDFINHGLAENKFFSDAFVPSK"
     repeat_region   1822. .1839
                     /rpt_type=inverted
     gene            1848. .2633
                     /gene="ddhA"
     CDS             1848. .2633
                     /gene="ddhA"
                     /note="DdhA; formerly known as rfbF"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose-1-P-cytidylyltransferase"
                     /protein_id="AAB49399.1"
                     /db_xref="GI:155496"
                     /translation="MEIQVKAVILAGGLGTRLSEETVVKPKPMVEIGGKPILWHIMKL
                     YSSYGINDFVICCGYKGYVIKEYFANYFMHMSDITFCMRDNEMIVHQKRVEPWNVTLV
                     DTGEDSMTGGRLRRVKDYVKDDEAFCFTYGDGVSDVNIAELIEFHKSHGKQATLTATY
                     PPGRFGALDIKDKQVRSFKEKPKGDGALINGGYFVLSPKVIDLIDGDKSTWEQEPLMT
                     LAAQGELMAFEHAGFWQPMDTLRDKIYLHELWEEGRAPWKVWE"
     gene            2536. .3711
                     /gene="ddhB"
     CDS             2536. .3711
                     /gene="ddhB"
                     /note="DdhB; formerly known as rfbG"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-glucose-4,6-dehydratase"
                     /protein_id="AAN23079.1"
                     /db_xref="GI:23321140"
                     /translation="MPVSGSQWIRCVTRFICMNYGKKAGHLGRYGNNKMINNSFWQGK
                     RVFVTGHTGFKGGWLSLWLQTMGATVKGYSLAPPTVPSLFETARVADGMQSEIGDIRD
                     QNKLLEAIREFQPEIVFHMAAQPLVRLSYSEPVETYSTNVMGTVYLLEAIRHVGGVKA
                     VVNITSDKCYDNKEWIWGYRENEAMGGYDPYSNSKGCAELVTSSYRNSFFNPANYGQH
                     GTAVATVRAGNVIGGGDWALDRIVPDILRAFEQSQPVIIRNPHAIRPWQHVLEPLSGY
                     LLLAQKLYTDGAEYAEGWNFGPNDADATPVKNIVEQMVKYWGEGASWQLDGNAHPHEA
                     HYLKLDCSKAKMQLGWHPRWTLNTTLEYIVGWHKNWLSGTDMHEYSITEINNYMNTK"
     gene            3729. .5042
                     /gene="ddhC"
     CDS             3729. .5042
                     /gene="ddhC"
                     /note="DdhC; formerly known as rfbH"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-4-keto-6-deoxy-D-glucose-3-dehydrase"
                     /protein_id="AAB49401.1"
                     /db_xref="GI:155498"
                     /translation="MSQEELRQQIAELVAQYAETAMAPKPFEAGKSVVPPSGKVIGTK
                     ELQLMVEASLDGWLTTGRFNDAFEKKLGEYLGVPYVLTTTSGSSANLLALTALTSPKL
                     GVRALKPGDEVITVAAGFPTTVNPTIQNGLIPVFVDVDIPTYNVNASLIEAAVSDKTK
                     AIMIAHTLGNLFDLAEVRRVADKYNLWLIEDCCDALGSTYDGKMAGTFGDIGTVSFYP
                     AHHITMGEGGAVFTQSAELKSIIESFRDWGRDCYCAPGCDNTCKKRFGQQLGSLPFGY
                     DHKYTYSHLGYNLKITDMQAACGLAQLERIEEFVEKRKANFKYLKDALQSCADFLELP
                     EATENSDPSWFGFPITLKEDSGVSRIDLVKFLDEAKVGTRLLFAGNLTRQPYFHDVKY
                     RVVGELTNTDRIMNQTFWIGIYPGLTHDHLDYVVSKFEEFFGLNF"
     repeat_region   5056. .5071
                     /rpt_type=inverted
     gene            5082. .5933
                     /gene="abe"
     CDS             5082. .5933
                     /gene="abe"
                     /note="Abe; formerly known as rfbJ"
                     /codon_start=1
                     /transl_table=11
                     /product="CDP-abequose synthase"
                     /protein_id="AAB49402.1"
                     /db_xref="GI:896326"
                     /translation="MRIVLTGGSGYIGSSLTPVLIKKYGRVYNIGRNTISEVSINGSK
                     EYCEFTYESLFDSLVELSPDLVINLAAGYYNDSGAPDLNVIDGNLKIPFIILEYFKSC
                     NYGRFINIGSYWEFSCSGRGVKGVNPYGIIKSTVRRLLDYYSKYNVIYTNLILYGSYG
                     DNDHRGKIVDCIIDAVNSNETLKLSPGEQKLNLVYIDDIIEAILYIVSSDNGQYDNET
                     LSIYTPTEHTVKEIVCFINEIKDNNLSLGGGRYRNDEVMAPDYKYRNIFHAKDKLKEY
                     ITSKIKK"
     misc_feature    6443. .6754
                     /note="part of non-functional tyv gene which has
                     accumulated numerous mutations"
     CDS             6755. .7375
                     /note="non-functional gene; orf6.8; remaining coding
                     sequence of mutated non-functional tyv gene"
                     /pseudo
                     /codon_start=1
                     /transl_table=11
     gene            7407. .8720
                     /gene="wzx"
     CDS             7407. .8720
```

```
                     /gene="wzx"
                     /note="putative O antigen export protein; Wzx; previously
                     called rfbX and orf7.4; TTG initiation codon"
                     /codon_start=1
                     /transl_table=11
                     /product="O-unit flippase-like protein"
                     /protein_id="AAB49403.1"
                     /db_xref="GI:409590"
                     /translation="MRVPTHIIVAASAWGSRLVSIFIQFYSIKILLDLLGTEGYAVFT
                     VIGSLVGWFLLADFGLGNSLQNQISYRRANHQEYQDLVLSAVIAIIPIFILFIILILT
                     LSPYISEFLLGGFDFLNNNQRSNIFKVASFIFLTTSIGNLAYKIWFSEHKGWVSNIIP
                     ALSSIVGLVFLMRLPSDGSNISEDIIFSIYCFYIPAAFFGVISTLFKVIPYLKCKNFL
                     NKLTLYTLIKNGGGFFLFSVLSALVLQVDYIVMSQTLVERDLVTYNIMSKTFGLINFI
                     YAALLQSLWPVCAEASSKLRFDNFYKIEKKYISFGFIIVIASSFVIFLLKDFIVNILA
```

Query Match 19.7%; Score 255; DB 1; Length 20125;
Best Local Similarity 51.2%; Pred. No. 8.9e-30;
Matches 653; Conservative 0; Mismatches 610; Indels 12; Gaps 2;

```
Qy         23 TTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTATCCAAAATAATAATTGCCGGGG 82
Db       7414 TACCTACACATATAATAGTTGCAGCAAGTGCATGGGGGAGTCGGTTAGTATCTATATTTA 7473

Qy         83 TGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCT 142
Db       7474 TCCAATTTTATAGTATTAAAATTTTATTGGATCTATTGGGTACGGAAGGATATGCTGTTT 7533

Qy        143 TTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAG 202
Db       7534 TTACCGTTATTGGTAGCCTAGTTGGTTGGTTCCTGCTTGCGGATTTTGGTCTCGGTAATA 7593

Qy        203 GACTGCAAAATTATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTA 262
Db       7594 GCCTTCAGAATCAAATATCATATAGGAGAGCAAATCATCAAGAATATCAAGATTTAGTTT 7653

Qy        263 AATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTT 322
Db       7654 TATCTGCAGTAATAGCTATAATACCTATTTTTATTTTATTCATAATTTTAATATTAACAC 7713

Qy        323 TTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAA 382
Db       7714 TGTCACCTTATATATCTGAGTTCCTCCTTGGTGGATTTGATTTTTTAAACAACAATCAGA 7773

Qy        383 CCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTG 442
Db       7774 GAAGCAATATATTTAAAGTGGCATCTTTTATTTTCTTAACTACATCCATTGGTAATTTAG 7833

Qy        443 CTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTAT 502
Db       7834 CATATAAAATATGGTTTTCAGAGCATAAAGGATGGGTTTCTAATATAATCCCCGCATTAT 7893

Qy        503 CTTATATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATC----TCAGTT 558
Db       7894 CATCTATTGTTGGACTCGTTTTTTTGATGAGACTACCATCTGATGGAAGCAATATTAGTG 7953

Qy        559 GACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTAT 618
Db       7954 AAGATATTATTTTTTCTATTTATTGTTTTTATATTCCAGCTGCATTTTTTGGTGTAATAT 8013

Qy        619 ATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCA 678
Db       8014 CTACTTTATTTAAAGTCATACCTTATTTAAAATGTAAAAATTTTTTAAATAAGTTGACAC 8073

Qy        679 ATTTTACGTAGATCTTCA--------GGGTTTTTTCTTTTTACTTTATTATCGATAGTGG 730
Db       8074 TTTATACTCTAATAAAAAATGGAGGCGGTTTTTTCCTATTTAGTGTACTATCGGCACTTG 8133

Qy        731 TGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAAT 790
Db       8134 TACTTCAAGTTGACTACATTGTTATGTCACAGACGTTAGTAGAAAGAGATTTAGTAACGT 8193

Qy        791 ATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAG 850
Db       8194 ATAACATAATGAGTAAAACTTTTGGTTTAATTAACTTCATATATGCGGCTTTATTGCAAT 8253

Qy        851 CATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAA 910
Db       8254 CACTTTGGCCTGTATGTGCAGAGGCAAGTTCAAAACTAAGGTTTGATAATTTTTATAAAA 8313

Qy        911 TGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTT 970
Db       8314 TAGAAAAAAAATATATATCGTTTGGTTTTATTATTGTTATTGCAAGTTCATTTGTTATTT 8373

Qy        971 ATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTA 1030
Db       8374 TTTTATTAAAAGACTTTATTGTAAATATACTCGCACCTGGAAAAGACTTCTATTTCCCTA 8433

Qy       1031 TTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATG 1090
Db       8434 TTAGTTTGATTATATTGTTTTCCTTTTATCAAGTGGTAAGGGTGTGGACTGATACATATG 8493

Qy       1091 CAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTACCACTACAAG 1150
Db       8494 CTATGTTTCTCATGAGTATTGGTAAATTAAAACCACTCTGGATTAGTGTTCCCTTTCAAG 8553

Qy       1151 CAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATCAGTGGAGTGC 1210
Db       8554 CTGTATTAAGTGGTTCACTTCAATGGGTTGGTGCAGTAAATTATGGTTTGGTAGGATTAC 8613

Qy       1211 TGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCACTAACTTACT 1270
Db       8614 TTTGTGGATTAATAGCATCCTTCTTAATTACCGTTTCATGGTGGTTGCCTTTCAGTTTTC 8673

Qy       1271 TAATTAAGGCAAATA 1285
Db       8674 GATCTACAGTAGATA 8688
```

RESULT 11
BX950851_16
WPCOMMENT
Sequence split into 51 fragments LOCUS BX950851 Accession BX950851

Fragment Name	Begin	End
BX950851_00	1	110000
BX950851_01	100001	210000
BX950851_02	200001	310000
BX950851_03	300001	410000
BX950851_04	400001	510000
BX950851_05	500001	610000
BX950851_06	600001	710000
BX950851_07	700001	810000
BX950851_08	800001	910000
BX950851_09	900001	1010000
BX950851_10	1000001	1110000
BX950851_11	1100001	1210000
BX950851_12	1200001	1310000
BX950851_13	1300001	1410000
BX950851_14	1400001	1510000
BX950851_15	1500001	1610000
BX950851_16	1600001	1710000
BX950851_17	1700001	1810000
BX950851_18	1800001	1910000
BX950851_19	1900001	2010000
BX950851_20	2000001	2110000
BX950851_21	2100001	2210000
BX950851_22	2200001	2310000
BX950851_23	2300001	2410000
BX950851_24	2400001	2510000
BX950851_25	2500001	2610000
BX950851_26	2600001	2710000
BX950851_27	2700001	2810000
BX950851_28	2800001	2910000
BX950851_29	2900001	3010000
BX950851_30	3000001	3110000
BX950851_31	3100001	3210000
BX950851_32	3200001	3310000

```
BX950851_33       3300001   3410000
BX950851_34       3400001   3510000
BX950851_35       3500001   3610000
BX950851_36       3600001   3710000
BX950851_37       3700001   3810000
BX950851_38       3800001   3910000
BX950851_39       3900001   4010000
BX950851_40       4000001   4110000
BX950851_41       4100001   4210000
BX950851_42       4200001   4310000
BX950851_43       4300001   4410000
BX950851_44       4400001   4510000
BX950851_45       4500001   4610000
BX950851_46       4600001   4710000
BX950851_47       4700001   4810000
BX950851_48       4800001   4910000
BX950851_49       4900001   5010000
BX950851_50       5000001   5064019
Continuation (17 of 51) of BX950851 from base 1600001 (BX950851 Erwinia carotovora subsp
```

```
  Query Match             10.8%;  Score 140;  DB 1;  Length 110000;
  Best Local Similarity   47.0%;  Pred. No. 1.3e-12;
  Matches  582;  Conservative    0;  Mismatches  635;  Indels   21;  Gaps    4;

Qy         52 TCATGGTTATCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATT 111
              ||||||  | ||    ||||||||     |  | |||| ||  ||||||    |     |  ||
Db      24748 TCATGGCTTTCTCGTATAATAATAATTGCCGCGCAGCTAATAAGTATAAGAATATTACTT 24807

Qy        112 TCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGG 171
                  |    ||  ||| |||    | |    |  |   | || |    | |||
Db      24808 AAACACTCTAGCTTGGATGAATTTGCTGCATATTCGCTTCTGGTTAGTATTGGTGGTTGG 24867

Qy        172 TGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGA 231
              | ||      | || || || || | |  |  |||||||||||||| || ||| |   |
Db      24868 TTTATATTAATGGACTTCGGGATTGCAATTGCTCTGCAAAATTATATTTCCGAACGGCTA 24927

Qy        232 GCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCT 291
              || || ||  |  | ||| | |     ||       ||| |  |  | |   || |
Db      24928 GCGAACAAAGAGCGATATCAAGATCTAATATCTAAGGCAATGTACTTTATTATTCCATGG 24987

Qy        292 ATTATTTTTTTTATTGCTTTATTTTATATTTTTCTGGGGTAATTTCCGCTAAATATCTT 351
                |||| | ||  | || ||| | |||||       ||  |  ||    |  ||
Db      24988 TACGCGGTTTTAGTCGCGATTTTATATCTATTTTCAACTAACATAGCATCTTTTTTGCTC 25047

Qy        352 TCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTG 411
              | | |  | |||       |  | |  ||| |    ||| | ||| |||    || | ||
Db      25048 CCGTTTAGTAATGTCAGTATTGATGTTAAAGCTGACATGATTTTTATTA---CAGGGCTA 25104

Qy        412 GTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTC 471
              ||||     | |  || || ||    | |   ||| |||||     | ||| |||
Db      25105 ATTTTTGCCACGACAGGGATTGGCAATGTATGTTACAAAATTTGGTATGCTGAAAATAAA 25164

Qy        472 GGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTA 531
              || |  | | | ||| |  |     ||| |       || || ||  |  |     | ||
Db      25165 GGATACATATCAAATATTCTGCCGGCAATTGGAGCAATAATCGGATTTTT---TCTCCTT 25221

Qy        532 TATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTT 591
              ||| |||     |   | ||    | |                 ||||  |  | |||  |
Db      25222 TATTTATTTATTGACCGCATGGAGAATAAATCTCTTGGCGGAATAATGTTATTTTATTCT 25281

Qy        592 CCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTT 651
              |||   |  | ||| | || | |  | |  | |  |   |  |         |  | ||
Db      25282 CCAACCGCGTTAATTGCTTTATTCGCTTTATTGAAAAAGACCGCACCCTGCATTTCTCTT 25341

Qy        652 AAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTT 711
                |     |||     ||   | | ||      ||| || ||  | ||| | | |  |
Db      25342 CCTAGGTTTGAATTATCTTTAGTCGTAAGAGGTAGTAAATTTTGGGTGTTATCTATCATG 25401

Qy        712 ACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACT 771
                  ||| || |   |||         ||||||  |  | || |||| |   ||  ||
Db      25402 GTGGCGTTAACGCTGCAGGT---------AGATTACGTTATTATGTCTCGATTCCTTTCT 25452

Qy        772 CCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATT 831
```

```
                ||| ||||||||     |||||  |||| | |  ||| || |||||  |  |  ||| |
Db      25453 GCTGAAGATATTGTCATCTATACGCTAACAACGCGAATATTCGGTTTTGGATACTTTCTA 25512

Qy        832 TATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAA 891
              |||  |  || |  |    ||||   |||||  |  | |  |||||  |   ||| ||  | |
Db      25513 TATTCATCTGTACTTGCAGCAACGTGGCCAACTTTTACTGAACTCTCAGTAAAGAAAGAT 25572

Qy        892 TGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTT 951
                |||| ||   |  ||  |          || ||||  || | |  |||  ||| ||
Db      25573 GCAGAAAAGATTTTAAGCATACTTAAGAAGTATGTTTTTTTTAGTTTGCTAGGTGTAGTG 25632

Qy        952 GGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGAT 1011
                | |||  |  |  ||  | |    |||||      |  |     || |  |   ||
Db      25633 ATTTTTACTTTCATATTTATTCTCACCAAAGATATCGTGGTAAGTATATTTACATCAGGA 25692

Qy       1012 ATTAATTATCAAGTTTCTAT------TTTATCTTTTATGTTAATTGGCATATATTTCTGT 1065
                |       |  | |  | ||       |  | | || | |||| ||| |  ||
Db      25693 AGCGGGAAGGTAATAACGATACCCAACTCCTTTATTCTATTAAATGGAACTTACCAGCTC 25752

Qy       1066 ATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATA 1125
                |||   | |||  ||     | | |||   | |||||||||  | ||||   |     |
Db      25753 GTTCTGATATGGACTGCATTATTTTCAACCATTTTGCAAAGTGTTAATTTAGTTCGCCTC 25812

Qy       1126 CTTTGGATATTAGTACCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCT 1185
                |   |||  |    ||  | || |||  |||| |||    |   ||||||    | ||
Db      25813 TTCCTGATTCTTACGCCTATTCAGGCAGCAATTAGTGCCTCACTCCAATGGGTGCTGGCT 25872

Qy       1186 AGTACGCTTGGAATCAGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTT 1245
                |  ||||  |   |||  |   | | ||  |  |  | || | |     || ||
Db      25873 GAGAATATTGGGCTGTATGGGATTACGATCGGGCTATTGGTTTCATATCTGGCAATTGCG 25932

Qy       1246 TTTTGGGGGCTTCCACTAACTTACTTAATTAAGGCAAA 1283
                |||||  ||| ||| |||  | | |||    |   |||
Db      25933 GTTTGGTTGCTGCCAATAAAATGCATAAAATATATAAA 25970
```

```
RESULT 12
AC116920
LOCUS       AC116920              105320 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 4415041-4520361 strain
            AX4, complete sequence.
ACCESSION   AC116920 AC116983
VERSION     AC116920.3  GI:42733646
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 105320)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 105320)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 105320)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 105320)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 105320)
  AUTHORS   Baumgart,C.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Feb 21, 2004 this sequence version replaced gi:28828137.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .105320
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="4415041-4520361"
     CDS             join(<84. .195,401. .445)
                     /note="GeneID exon scores (in order of location ranges):
                     5.39, -2.56 - GSCJ_ID dd_03376"
                     /codon_start=2
                     /product="hypothetical protein"
                     /protein_id="AAS38624.1"
                     /db_xref="GI:42733679"
                     /translation="IEILASSTKHFQSICRRLYRIFAHAFYHHREAYNEYENSLLPKS
                     AIIIKEN"
     CDS             join(2017. .2175,2265. .2273)
                     /note="GeneID exon scores (in order of location ranges):
                     16.22, -1.64 - GSCJ_ID dd_00543"
                     /codon_start=1
                     /product="similar to Homology to rat S29; Rps29bp
                     [Saccharomyces cerevisiae]"
                     /protein_id="AAS38610.1"
                     /db_xref="GI:42733647"
                     /translation="MARELWLTHPRNFGPGSRTCRKCGNHHGIIRKYDLNMCRRCFRT
                     DAEAIGFNKYR"
     CDS             complement(2879. .3808)
                     /note="GeneID exon scores (in order of location ranges):
                     24.99 - GSCJ_ID dd_00545"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38611.1"
                     /db_xref="GI:42733648"
                     /translation="MVEQCPLDMCLSYNLCTSNSSFGNLYFGCKDEYGNKVQTKNGIR
                     CFSHFKPILLIPILVFHLIFALVLSLIKHNSFRIPISQIIIFLQINSVVFFSYSGIYI
                     LPLFGMSVDMLDGYCLWEGLNYNFKSFLSFSIIISLFLIGSTDITPIVLIKIYEKIPL
                     IANNNSIRNRKKNINDSEKTIIEKTWCPRFLKEKIKKRINEKSNSKVYYWKSIWSFIS
                     TFCNTFDVQLNIFNTILYCYYTSIQIVYMSFYFFLATSSNNSSRDLNNYINLAQLIIF
                     IVANSNLLQLGSLFGQGLVLTSVSFSFSLVLIN"
     CDS             complement(join(4065. .4273,4441. .4587,5023. .5188))
                     /note="GeneID exon scores (in order of location ranges):
                     8.11, 1.77, -2.21 - GSCJ_ID dd_00546"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38612.1"
                     /db_xref="GI:42733649"
                     /translation="MIISENASLIIFNSIKLNENVGSFLKSSNDKSLHINTLIYNDNK
                     IPYDYLFSIENYLIASFNIAKFGGVFQFGGAYSNKNDAFINPISKFINNKFSSNIGSQ
                     QGELQKAIFSNDSINVINGVDTLVKFYFFDYFNNTFASLSMPLNISITRLLDSSTKFS
                     LSIMAGKGIPLSV"
     CDS             complement(join(5249. .6306,6524. .6773))
                     /note="GeneID exon scores (in order of location ranges):
                     67.43, 19.81 - GSCJ_ID dd_00548"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38613.1"
                     /db_xref="GI:42733650"
                     /translation="MVFVNSTFRFTVFTGELNATNSLFRFIGSSFINGGIVFNLINSD
                     SSSGDSSSNSSSNSSSNSSSSSNSNIICISGCGVSFSDNNTTITFSKSVDPLIIIIAY
                     RYSYVSIENVKIKGYRSSNPMVISSEGSEIIMNSIFIEDCISNSLISTRKGGFSDGGS
                     IEINKIKFKDNTFDSIIDVTGVSNVTISELITTCEMNNKFYSMIVMATADDVYDVNNQ
                     YLSKSNVSFSFMELSCIGSFQFERALVELLIVNMSASFNQFLFTLTKESILVIDRCDF
                     ILDPLTCSTKSTAFLVENISDLYIVDSIFGYSGTGIVVSTNGSSITLVNTRILSIILS
                     VPVFTIGHSTLDFYNCTFEQTYIPSTSLIDSFISNITIDICTIDTLGDVFLDGFGSNI
                     IINQLVMTNFFTQYRLFQSNSDSIIISNSLFINIHGFGYQLFQ"
     CDS             complement(join(6894. .7206,7397. .7407))
                     /note="GeneID exon scores (in order of location ranges):
                     4.13, -2.65 - GSCJ_ID dd_00551"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38614.1"
                     /db_xref="GI:42733651"
                     /translation="MKKFNNESNGTIINPFQDLCILNNILQQSFYNFNNNYYENINFN
                     QNNLNSIIKIYIGYGNYNLTCQSISISNFNNISISPYNFNQNQTIIENLFLNLENTNL
                     IINGP"
     CDS             complement(join(7649. .9423,9499. .12796))
                     /note="GeneID exon scores (in order of location ranges):
                     89.47, 176.06 - GSCJ_ID dd_00554"
                     /codon_start=1
                     /product="similar to Plasmodium falciparum. Hypothetical
                     protein"
                     /protein_id="AAS38615.1"
                     /db_xref="GI:42733652"
                     /translation="MKIYKYSLTFIFLIFIFLNNVCSQNINNNNNNINQYEDNNIINI
                     YINSNNQNNGNGTIINPYNSICQIESILNNINNNNNNNNNKNNNNIATSIIIVNIGY
                     GIYNDTNCYQFLILNCKNVTFKPYISNSSSSKDNNNNNNNNNNNNNNNNNNNNNNNN
                     KEEEEGEQKVSIYNSFKIINSKFYIQGPFITFYGSWNISNSEMNIIDNAYIIFKDKYN
                     IYSIISNSVIKLINSTMISSIYDQKYMIEDSLLYFQKSTLKQVSINFNSNPKFASSGG
                     GGRVEIIDSKVMGSFFQMGFSDINVKNSIWQETVLFTVADNSKFENILFNQNDLKLWS
                     EYYNSVGIINIDKELFNENGSKITILNSEFRANPKYGALSLIVSLDIEQDVLIILYNI
                     RMIGFHKYESMFASTSYGGNITISLELVEFDNCVSPSMFTMENTNLVISDLIFTNCIF
                     ETMLEARNSSNIEISNLSINSTQNNIIQKFLSMPKDENSFEMSSSIEFSNIFLESIEG
                     SFEFRNSIVNFEFVTMISIGSDSFINGYQGSIVNIISSDFISNPFGKINQLFYFEQSS
                     ITISQSNFFDSAVPIVMAVKNSSVLIDNNSKFIDLQTSQPIITVADDSILEIYSTLFN
                     NCRISSSSLIQSESSLVYISETNFLLVSQQLFNSSRSKISMNIVNVSQSFTQSQLITS
                     VDDYIFIASSVFDGNHGLSYQMFTLNGSILNITQTQFNGNSFNPLIKSVNSTIEFVET
                     LYSNNTGSFINSKDDQLISISFLEFYDSRLPINYLLSSDSTKTIILNQIFINNNYFTQ
                     FFISSFNAETFEIDTFEFGNNLIVHVPMLLCPYNQFDGLGGFISLTQVSNVLIKNFNI
                     RNNQIGTTFLTFYDCVFSKYNSNNSNSNSYSNSYSNTGSGVGGGSSGSCSGSSSGSSS
                     GSSSCSGSVNIQNSQFTSNTIQSDISLYSESNYLIESKSSEVSISLTYFNGNQGNGKG
                     LILVINSNITIYGCSIDSNFFENGTGAILSFFPDYLIDDSFQWFVIVQSLLTTNFGIY
                     GGVLYFNSPQSLNNPLSKIENNSFINNRALQQGGVVYTTDSPFNENMTKVLNSNFFIG
                     NTAFCGPNFSSNMMKNPQLSNPISQVVDDGSSIIEYQFVDYYGNLFTPFSSILTIYVN
                     RTSTNHLMNVTNNVNAGMGKFTFEFIGSINDEYLIGVNSPTPVLAKVYVDGCSPYQFY
                     NTTSEMCTYCPNNYALSQLLGKCYPCNTNIMNCLASTVFAIDGYYLVNNDVSMVEQCP
                     LDMCLSYNLCTSNSSFGNLCFSCKDEYGNLVQTKNGIRCCSHFKPILLIPILVFHLIF
                     ALVLSLIKHNSFRIPISQIIIFLQINSVVFFSYSGIYILPLFRMSVDMLDGYCLWEGL
                     NYNFKSFLSFSIIISLFLIGSTDITPSIIIKIYKKISTRKDNRNLSLLINDRNITPLQ
                     QQQQPIAKYPNYIIKILEGKIKHRSIYYWKSIWSLFQLFVIPLAFNAMSLVISKQVGS
                     ENYLAVDFSIDFQSNSNQACILIAILIFIILAIVLFMQLILLLSKFKIGGNRIFKNRI
                     TESLLKIQQYQLEYRTNFKWFPIVINSRSIIICVLSIVLLFKAQYYIPIILLIQIFYT
                     IISLLINPIKWKYLRDFNNIINLLQVVILITANSNLLQSKSLFNEGLIITIITFSFAL
                     LSIISIFLLYNGKNKKIRRKLK"
     CDS             join(13910. .13922,14046. .14224)
                     /note="GeneID exon scores (in order of location ranges):
                     2.36, 2.61 - GSCJ_ID dd_00558"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38616.1"
                     /db_xref="GI:42733653"
                     /translation="MPILGSISTFSQFNVKKSKNDSISSFLKSNVGYSNSMVSEIDGR
                     FDNLPIIIQKPPNKPNAYI"
     CDS             complement(join(14483. .19185,19254. .19814,19913. .19922))
                     /note="GeneID exon scores (in order of location ranges):
                     556.18, 66.59, 0.22 - GSCJ_ID dd_00562"
                     /codon_start=1
                     /product="similar to unknown protein; protein id:
                     At1g65440.1 [Arabidopsis thaliana]"
                     /protein_id="AAO50823.2"
                     /db_xref="GI:42733654"
                     /translation="MSSKNYSDDEIENEENIHPKKKFKNLNKNKDDESSELEEYHDSS
                     EEDENEPNQYQYDDFVVGGDDDERRKRKKKEKRKKSSSNSDNEDDNEEEEEEDRGGGG
                     GSGSGGSGGSGGRNSSSPKRTNNHLDEDIDGDEDGDDGSDGSESESESEEDEEVAQAR
```

```
                     RERRRKRREERRKREKGGLQRLKKLRASNKDLSDKEDDVEDGYNGEQEDDGFVVDSHG
                     KPVKRSHERERMSYESSNRSRIMSDIFDDDEEEGYGSGEEEEDSGINKTEADRAKDRL
                     AAIREQYEPSLLEEKHFTDADEEIRNKNVPERLQSRKGTQYAGEMACEEEAEWIYEVA
                     FEQRDFQQQQQRNPNTEPVRDAKAITAIFKILQFIQRDLLEIPFIYTYEKDIYEPYFT
                     LQELWNIFDLDEKWAHMKVNKRNLEQIGSNNQQLEPYKSVLLESRSEESISDLYDLFQ
                     MINGIQKNTNLLNGSSSLGSGGDGYDDKEGGSSATQGNTVGGLSRQKKAIKRDLYTIY
                     TKAGLSKFLPNFGMTAQEFGQNLMDNYTTNKPRDQMAEPSTMALVHICLEADNKDRVL
                     QATRYMMAQEIGYDPHVRYSVRMIYRKYAHITTTPTLKGFKEIDVFHPYFTVKSIQEK
                     PAHLFEDSQYLLILKAEKEGFIKSTMAITEKTHNSVIIPEMESLYLSDGTSAITQQWN
                     EQRKLIIREALNKFLYPVLEKELRNKLLTEASNRVAFECAKKLEDKIRVAPWRPVSSS
                     SGIGHSNSSVLFGGSRANNGYSNDDDEDEDDDEEVDYDRDGNPIPFKILSLCWGSDKI
                     PTMGAVLNSDGEVLSHIKLDFICDRLGESLKEKKEKDIKRLEEICQEYQPRLVLVSAT
                     EMDSKRLYEEVKDHLQRWSNGERRIIRKSVLLNYYSPEIGLSLQTSSRLQEEFKEYPP
                     ILRHTIAVGRCALDPITEYASLCNDHNEILFLKLHPLQDMIGKDYLVKLLHRCFINVV
                     NAVGVDINRMIQCRFTSATLQFVSGLGSRKAQMLLNSLFRRGGPVTSRQSLEKVLSQD
                     VIYRNCIGFIQIRERHAADYKADLLDDTRIHPTSYPITYRIAAEALDKNLDERNFHSY
                     IEDIIKRPKKLDRLDLDGFADIIESHEDRPARKLLYFIKKELTNPFADIRHPYEEPSA
                     DTIFEWLTGETNQSLRRGTLVTVTTIRTFDNSVKCRLDNGLEGSIPTDCISDNGETKS
                     LGRGVTINCRVMGVDKYQFTVSLSCKPSDLNPSFWEETIFRELKENGANQYLRLEEVA

  Query Match             6.6%;  Score 85.2;  DB 2;  Length 105320;
  Best Local Similarity   44.2%;  Pred. No. 0.00026;
  Matches  392;  Conservative    0;  Mismatches  493;  Indels    1;  Gaps    1;

Qy       215 ATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTAC 274
Db     20433 AAAAAATAAAAAAAAGATAAAAAAAGAGAAATCATTTTTCTTTTTTTATTTTACTTTTC 20492

Qy       275 ATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAA 334
Db     20493 TTTTTATTTTTTATTTTTTTTTTTATTTTTCTTTTTATTTTTTTTTATTTTTTATTTTTTT 20552

Qy       335 TTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCT 394
Db     20553 TTTATTTTTTTCTTTTTATTTTATTTTTTTTTTTTTTATCGTTCAGCACCGAATCGATAT 20612

Qy       395 TTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATAC 454
Db     20613 TTTTAACCCCACTTTGGAATTGAAATAATTTTTTTTTTTCTTTTATCATTATTACCGTTT 20672

Qy       455 TTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAG 514
Db     20673 TGGTTTTGGATTTCCAATTTTTTCCTTTTTCTTTTTATTTTTCCATTTTTTTTTTTAAAA 20732

Qy       515 GTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCAC 574
Db     20733 AAAATATTATTTTCCTTTATTTTTAATTAAAGTGTATTTATTTAGAATTTAAATTAAAAC 20792

Qy       575 TAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACA 634
Db     20793 TAAAACTAGTTTTTATTTTAATTTATTTTTAATAATTAAATAAAATTTTA-ACTTATCTT 20851

Qy       635 TAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTT 694
Db     20852 TCAAAATATTTATTTATTTTTAAAAAAAAGAGAGGGAGACTTGTGAACTTCCAATTTTTT 20911

Qy       695 CAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCA 754
Db     20912 CCTTTTTCTTTTTTTTCTTTTTTTTTTTGCTTTTTTTTTTCTTTTCAAATTTTTTTTTTA 20971

Qy       755 TTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTG 814
Db     20972 GTTTTTTCAAAGTAACAATTTGTTCGTGTGGTCAATTCTAAAAAAAGGTAAAAAACAGTG 21031

Qy       815 GTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAAT 874
Db     21032 TTTAAAACAATAGTTTATAACTTGTTAATTTGGTGTTACAATAATCTTTAAAATTTTTAA 21091

Qy       875 TGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTG 934
Db     21092 AATTTTAAAATAAAATTTTTTTTTAATAAAGACGACCATAATTGGGGGTGGAGGGGGGAA 21151

Qy       935 GCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTT 994

                                                                                
Db     21152 ATTAATTATAATAATTTTTGTGATAAAATTTGGATATTATCTATTTATATCACTTTAGGAA 21211

Qy       995 CAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCA 1054
Db     21212 AAGAAATAAAAATAAAATAAAAATAAACTAAAACTAGTTAAACAATTTTTTTTTTTTTG 21271

Qy      1055 TATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATT 1100
Db     21272 TTTATTTCTTTTTATTTTTTTTTTTTTATTTTTTTCCATATTATT 21317


RESULT 13
AC011718
LOCUS       AC011718               94384 bp    DNA     linear   PRI 09-APR-2003
DEFINITION  Homo sapiens chromosome 22 clone b2280111 map 22q11, complete
            sequence.
ACCESSION   AC011718
VERSION     AC011718.2  GI:6598880
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 94384)
  AUTHORS   Hu,P., Dumanski,J. and Roe,B.A.
  TITLE     Homo sapiens Chromosome 22q11 BAC Clone b2280111 In DGCR Region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 94384)
  AUTHORS   Hu,P., Dumanski,J. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 94384)
  AUTHORS   Hu,P., Dumanski,J. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 94384)
  AUTHORS   Hu,P., Dumanski,J. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 94384)
  AUTHORS   Hu,P., Dumanski,J. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Dec 17, 1999 this sequence version replaced gi:6041881.
            Because these overlapping clones came from different libraries.
FEATURES             Location/Qualifiers
     source          1. .94384
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11"
                     /clone="b2280111"
ORIGIN

  Query Match             6.6%;  Score 85;  DB 8;  Length 94384;
  Best Local Similarity   45.2%;  Pred. No. 0.00029;
  Matches  393;  Conservative    0;  Mismatches  470;  Indels    6;  Gaps    2;

Qy       284 TTATAGCTATTATTTTTTTTATTGCTTTATTTTATAT-TTTTTCTGGGGTAATTTCCGCT 342
Db     80804 TAATATATAATATATGATATATAATGTTATATTACATATAATTATATTATATTATATTAT 80863
```

```
Qy        343 AAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACC 402
Db      80864 ATATATCATAATATATTATATAATATATAATATATATCATAATATATTATATAATATATA 80923

Qy        403 TCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCC 462
Db      80924 ATATATATCATAATATATTATGTAATATATATAATATATCATTTTATATATTATATATAA 80983

Qy        463 GAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTC 522
Db      80984 TATGTCATAATATATTATATAATATTATATATAATATATCATAATATATTATATATTATA 81043

Qy        523 GGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTC 582
Db      81044 TTATTATAAAATATATAATACTATATAATGTTATATATAATATAATATTATATTAGATTA 81103

Qy        583 CTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTT 642
Db      81104 TAATATATAATATTATATTAGATTATAATATATAATATTAGATTATAATATATAATATAT 81163

Qy        643 TATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTT 702
Db      81164 TAGATTATAATATACATTATATTATAATATATTATTGTATAATATGTTATAATATAATAT 81223

Qy        703 TTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAA 762
Db      81224 ATATTATATATTATATTATACATAATATTATATTATATATAATATATAATATTTTATA 81283

Qy        763 AGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTC 822
Db      81284 CAACATATTATATATAATATCTTATATAATATATTATACAATATATTATATTGTATAATA 81343

Qy        823 TTTTTTATTTA-----TACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGA 877
Db      81344 TATTATATATATTATGTATTATTATCTTGTATATATAATTATATATAATTATATAATTAT 81403

Qy        878 GAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCT 937
Db      81404 ATATATTATATTATAGGTAATAATTATATATAATTATATATTATATTATATGTAATAATT 81463

Qy        938 CACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAG 997
Db      81464 ATATATAAGTATATATTATATTATATGTAATAATTATATATAATTTATGTATTATATTAT 81523

Qy        998 TAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATAT 1057
Db      81524 ATGTAATAAATACACATAATTATATATATTTTATTATATATATTATGTATAATTGTATAT 81583

Qy       1058 ATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATT 1117
Db      81584 TATATTATATATAATAATTATAGGTTATATTAGATATATTAATTGTATTTTATATTATTT 81643

Qy       1118 TAAAAATACTTTGGATATTAGTACCACTA 1146
Db      81644 TTAATATTAGTTAATATTAATTAATAATA 81672
```

```
RESULT 14
AC091214
LOCUS       AC091214               143331 bp    DNA     linear   PRI 07-MAR-2002
DEFINITION  Homo sapiens 12q BAC RP11-13F9 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC091214
VERSION     AC091214.20  GI:14786798
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 143331)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
            Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
            Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
            Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
            Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
            Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
            Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F.,
            Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
            Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,
            Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
            Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
            Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
            Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
            Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
            Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B.,
            Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
            Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
            Washington,C., Watlington,S., Williams,G., Williamson,A.,
            Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
            Zorrilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143331)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 143331)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 143331)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 143331)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   6  (bases 1 to 143331)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   7  (bases 1 to 143331)
  AUTHORS   Worley,K.C.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   8  (bases 1 to 143331)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 17, 2001 this sequence version replaced gi:14578078.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

            QUALSTAT-REPORT.
FEATURES             Location/Qualifiers
     source          1. .143331
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-13F9"
     misc_feature    complement(1. .2005)
                     /note="overlaps bases 1. .2005 of clone AC079409"
                     /function="clone overlap"
     repeat_region   511. .703
                     /rpt_family="LTR16A"
     repeat_region   complement(825. .1128)
                     /rpt_family="AluY"
     repeat_region   complement(3191. .3292)
                     /rpt_family="L2"
     repeat_region   3856. .3887
                     /rpt_family="AT_rich"
     repeat_region   complement(4061. .4369)
                     /rpt_family="LTR33A"
     repeat_region   4457. .5350
                     /rpt_family="L1M1"
     repeat_region   5351. .5373
                     /rpt_family="T-rich"
     repeat_region   complement(5374. .6088)
                     /rpt_family="L1PA7"
     repeat_region   6109. .6146
                     /rpt_family="(TTTTA)n"
     STS             7538. .7716
                     /standard_name="61535"
     repeat_region   12648. .13327
                     /rpt_family="L1M1"
     repeat_region   13353. .13387
                     /rpt_family="AT_rich"
     repeat_region   13423. .15610
                     /rpt_family="L1M1"
     repeat_region   15611. .15710
                     /rpt_family="FLAM_C"
     repeat_region   15718. .15889
                     /rpt_family="(TA)n"
     repeat_region   15909. .16212
                     /rpt_family="(TA)n"
     repeat_region   complement(16213. .16263)
                     /rpt_family="FLAM"
     misc_feature    16261. .17578
                     /function="unresolved tandem repeat"
     repeat_region   16264. .16399
                     /rpt_family="(TA)n"
     repeat_region   16405. .16446
                     /rpt_family="AT_rich"
     repeat_region   16447. .17315
                     /rpt_family="(TA)n"
     repeat_region   17336. .17504
                     /rpt_family="(TA)n"
     repeat_region   17599. .17636
                     /rpt_family="(TA)n"
     repeat_region   17648. .17764
                     /rpt_family="AluJo/FRAM"
     repeat_region   17774. .17807
                     /rpt_family="AT_rich"
     repeat_region   17811. .18491
                     /rpt_family="L1"
     repeat_region   18494. .18612
                     /rpt_family="AluJb"

  Query Match             6.5%;  Score 84.2;  DB 8;  Length 143331;
  Best Local Similarity   44.4%;  Pred. No. 0.00033;
  Matches  562;  Conservative    0;  Mismatches  693;  Indels   12;  Gaps    5;

Qy         32 ATTTAATTGTTGCAGGTTCATCATGGTTATCCAAAATAATAATTGCCGGGGTGCAGTTAG 91
              |||| ||   || |  || ||   |  |        |  ||  |  || || ||  | |
Db      16203 ATTTTATATATGTATTTTAATAGAGAATGGTTCTCACTATGTTGCCCAGGCTGGTCTCAA 16262

Qy         92 CAAGTATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGT 151
                | ||| |      |||| | |    ||  |  | |  | |||   || | || ||
Db      16263 AATATATATATATATTTATATATTATTTATATTTATATTATTATATTATTTATATTTACA 16322

Qy        152 TAACTGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAA 211
              |   |  |   |    ||    ||     | || || ||   |||  || |        |
Db      16323 TTTTTTATAATTATAAATATATTATTTATATTTATATTTATATATTTATATATTATTTAT 16382

Qy        212 ATTATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCAT 271
              ||||||  | | |    |  | | | |  ||  |||     |||||  |  ||   |
Db      16383 ATTATAATTTATATATAATTTCTATATAATTAATATATTTGATATATAATTTATATTAAA 16442

Qy        272 TACATCTAAGCTTTATAGCTATTATTT-TTTTTATTGCTTTATTTTATATTTTTTCTGGG 330
              |   | ||    |  ||  ||| |||| ||| ||||  | |||| |||||||  |  |
Db      16443 TTATTATATATATATTATATATAATTTATTTATATTTATATATTATATATTTATATTAAT 16502

Qy        331 GTAATTTCCGCTAAATATCTTTCTTCTTTTCATG---AGGTATTACAGGACAAAACCAGA 387
               ||| |    || |||| |   |  |||       |  ||||| |       |
Db      16503 ATAAATATATTTATATATATAAATATATTTATATTACATATATTATATATATTTATATAT 16562

Qy        388 ATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTAT 447
              ||    | |||| | | | | | | | | | | |    ||       ||   |
Db      16563 ATAAATATATTTATATTACATATATTATATATATTTTTATATATAAATATATTTATATTA 16622
```

```
Qy        448 AAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTAT 507
Db      16623 CATATATTATATCTATTTATATATTTTATGTATATTTATATATTATATATATTTGTATAT 16682

Qy        508 ATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAA 567
Db      16683 ATTATATATATTTTATATATATTATATATATATTTTATATGTATTTATATTTTTATATAT 16742

Qy        568 TTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATAT 627
Db      16743 ATTTTATATATCTTTATATATATCTTATGTATATTTATATATATTTTATATATTTATATA 16802

Qy        628 AGATACATAAAGCTTTATCATGT-TAAAACAACAAAATCTCATTATATAGCAATTTTACG 686
Db      16803 TTATATATTTGTATATAATATATATTTATATATATTTTATATATATTTATATAGATTATA 16862

Qy        687 TAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTA 746
Db      16863 TATATTTTTATATATTTTATATATTTATATATTATATATATTTATACATATTTTATACAT 16922

Qy        747 TATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAA 806
Db      16923 TCATATATTATATATATTTATATATATTTTATATATTTATATATTATATATTTATATATA 16982

Qy        807 AATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATG 866
Db      16983 TTTATATATTTATATATTGTATATATTTATATATATATTTATATATTTATATATTGTATA 17042

Qy        867 TGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATAT 926
Db      17043 TATTTATATATATTTTATATATTGATATATATTATATATTTATAT------ATTTTATAT 17096

Qy        927 TTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACA 986
Db      17097 ATTGATATATATTATATATATTCATATATATTTTATATATTTATATATATTATATATTTA 17156

Qy        987 GATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTT 1046
Db      17157 TATATATTTTATATATTTATATATTATATATATTTATATATATTTTATATATTTATATAT 17216

Qy       1047 AATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAG 1106
Db      17217 TATACATTTATATAT-TATATATATTTATATATTATATATATTTATATATTATATATTTA 17275

Qy       1107 TATGAATTATTTAAAAATACTTTGGATATTAGTACCACTACAAGCAATAATTGGTGGAAT 1166
Db      17276 TATATTATATATATTTATATATTATATATATTTATATATAGTTTATATATTTATATATTT 17335

Qy       1167 AGCACAATGGTATTTTTCTAGTACGCTTGGAATCAGTGGAGTGCTGCTTGGCTTGATTAT 1226
Db      17336 TATATATTTATATATTTTTATATATATTTATATATTTTTATATATTTTTATATTGTTTAT 17395

Qy       1227 ATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCACTAACTTACTTAATTAAGGCAAATAA 1286
Db      17396 ATATTTATGTATTTATATATTTATATATATTTTATATATTTATATTATATATTTATATAT 17455

Qy       1287 GGGATAA 1293
Db      17456 TATATAA 17462
```

```
RESULT 15
AC125567
LOCUS       AC125567               170627 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-9H22, *** SEQUENCING IN PROGRESS ***,
            4 unordered pieces.
ACCESSION   AC125567
VERSION     AC125567.5  GI:24817949
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 170627)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170627)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 170627)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:23096537.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
```

```
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GDDV
              Center clone name: CH230-9H22
          -------------- Summary Statistics
              Assembly program: Phrap; version 0.990329
              Consensus quality: 162619 bases at least Q40
              Consensus quality: 164392 bases at least Q30
              Consensus quality: 165269 bases at least Q20
              Estimated insert size: 169231; sum-of-contigs estimation
              Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *         1   109158: contig of 109158 bp in length
          *    109159   109258: gap of unknown length
          *    109259   110280: contig of 1022 bp in length
          *    110281   110380: gap of unknown length
          *    110381   111540: contig of 1160 bp in length
          *    111541   111640: gap of unknown length
          *    111641   170627: contig of 58987 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170627
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9H22"
     misc_feature    1. .1284
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    3636. .4924
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH341988"
     gap             109159. .109258
                     /estimated_length=unknown
     gap             110281. .110380
                     /estimated_length=unknown
     gap             111541. .111640
                     /estimated_length=unknown
ORIGIN

  Query Match             6.5%;  Score 83.8;  DB 14;  Length 170627;
  Best Local Similarity   45.1%;  Pred. No. 0.00036;
  Matches  405;  Conservative    0;  Mismatches  482;  Indels   12;  Gaps    2;

Qy        242 AAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTT 301
              | | |||| |    | |||||||| |   ||  || | |  ||| |   |||  ||||||
Db      28085 ATATTTATAAAATTTTTATTAAATTATTTTTTTATTTTATTTTTTTTTTTATATTTTTT 28144

Qy        302 TTATTGCTTTATTTTATATTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTC 361
              |||||  ||| |||| | |||||  |      || |     |  | | ||| || ||||
Db      28145 TTATTATTTTTTTTTTTTTTTTTATTATTAATATATTTTTTTTTTTTTTTTTTTTTTT 28204

Qy        362 ATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTT 421
              ||    | ||       | |   |   |  | |||      |  | | |  |||| | ||
Db      28205 ATTTTTTTTTTTTTATTATATTTATTTTAATTTTTAAATTTTATTTTATATTTTTAATTT 28264

Qy        422 CTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAG 481
              ||||     |   | | |  || ||      |||||       ||  |    |  ||
Db      28265 TTATTTTTTTTTTATATTTAAATTTTATTTATATTTTTTTTTTTTTTATTATTTTTTAATT 28324

Qy        482 CTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACT 541
              | | | |||     |||| ||  | ||      |    |  |||   || | |
Db      28325 TTTTTTTTTATTTTTTTATTTTTATTTATTTAATATTTATTTATAATATTATTAAATT 28384

Qy        542 ATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTA 601
              |       | |   |    ||    ||||   |  |  |  | ||| ||    |  |
Db      28385 TTTTTTTATTTTTTTTAATTATTTTTTTATTTTTTTTTTTTTTTTATTTTTATTTTTATT 28444

Qy        602 TGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCA-----TGTTAAAAC 656
              |  |     || |  | | || | | ||  |   |      ||| | |     | || |
Db      28445 TTTTAATTTTTTTATTTTTTTTTTTTTATTTTATTTTTTTTTTTTTTATAATTTTTTTATTT 28504

Qy        657 AACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTT 716
              ||  | | |   ||| |    |||||  || || ||    |||| | ||||    ||
Db      28505 AAATTATTATTTATATTTTATTTTTTTATTTATATATTTTTATTTTTATTTTTTATAATT 28564

Qy        717 ATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCA-------TTTCTCAAAGGCTAA 769
              || |  ||  |  |  |   || |  || | |  | |       ||| |   |   |
Db      28565 TTTTTTTATTTTTTTTATATTAATATTTTTTTTATTAATTATTTTTTTATTTTATTTTTT 28624

Qy        770 CTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTA 829
              |  |  |  |||| ||   | || | |      ||      |||||  |||| | ||||||
Db      28625 TTTTTATTTTTATTTTTTTTTTTTAAATTTTTTATATTTTTTTTTTTTTTTATTTTTTTTT 28684

Qy        830 TTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGC 889
              |   |      |||| |       |||     | ||| |  |    ||      |  |
Db      28685 TAATTTTAAATATTATTTTTTTTTTTAATTTTTTTATTTTTTTTTTTTTTTTTTTATATTT 28744

Qy        890 AATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTG 949
                          ||          |||       | | || |  ||     | ||  | ||
Db      28745 TTATTTTTTTTTTTTTTTTTTATATATTATATATTTTTATTTTTTTAATAATAATTTTTT 28804

Qy        950 TTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAG 1009
              |   ||    ||  ||| |||||||  ||  | |  | ||  ||| |  | |      |
Db      28805 TATAATAATTAAATTTTTTTTATTTTATTTTAAATAATATTATTTAAAAATTTTATTATT 28864

Qy       1010 ATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTC 1069
              ||  || | |  |  ||| | |||| | ||||||  ||  ||   || |||||   ||||
Db      28865 ATTATATTAAAAATATTTTTTTTTTTTTTTTTTATTTTTTTTATTATTTATTTAAATATTT 28924

Qy       1070 GCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTT 1128
                | |  | |    ||||  | | |||||    | |   | |||| ||   ||| ||
Db      28925 TATTATTATATTTTTTTTATTTATTATTATTTTTTATTTAAATTTATATATTTATATTT 28983


Search completed: May 11, 2006, 19:15:03
Job time : 4661.07 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 07:09:59 ; Search time 589.159 Seconds
(without alignments)
14626.705 Million cell updates/sec

Title: US-09-423-093B-4_COPY_12762_14054
Perfect score: 1293
Sequence: 1 gtgaaagttcaattgttaaa.....ttaaggcaaataagggataa 1293

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1293	100.0	22080	2	AAX06751	Aax06751 Salmonell
	2	258.2	20.0	14309	14	AEC31432	Aec31432 E. coli t
	3	81.2	6.3	7597	6	ABL33013	Abl33013 Human imm
c	4	74.8	5.8	158001	12	ADL17884	Adl17884 Human pho
	5	73.8	5.7	842	14	ADZ71010	Adz71010 Human chr
c	6	72.4	5.6	1524	14	ADZ70897	Adz70897 Human mat
c	7	72.4	5.6	4660	14	ADZ70902	Adz70902 Human mat
	8	72	5.6	2131	14	ADZ71009	Adz71009 Human chr
	9	71.8	5.6	13511	6	ABL32281	Abl32281 Human imm
	10	71	5.5	15548	6	ABL34155	Abl34155 Human imm
	11	71	5.5	61020	4	AAS46788	Aas46788 Tumour su
c	12	69.4	5.4	828	14	ADZ70929	Adz70929 Human chr
c	13	69	5.3	220	8	AAL60729	Aal60729 Yersinia
	14	69	5.3	5449	6	ABL33641	Abl33641 Human imm
	15	68.2	5.3	115218	8	ACA64845	Aca64845 Human HNR
	16	68	5.3	3683	8	ABZ10199	Abz10199 Haematopo
	17	67.8	5.2	6292	4	AAS46735	Aas46735 Tumour su
	18	67.6	5.2	7728	6	ABL32077	Abl32077 Human imm
	19	67.6	5.2	7728	6	AAD28367	Aad28367 Human che
c	20	67	5.2	1039	14	ACL64817	Acl64817 M. xanthu
	21	66.6	5.2	745	14	ADZ70917	Adz70917 Human chr
c	22	65.6	5.1	1554	14	ADZ70999	Adz70999 Human chr
	23	65	5.0	7990	6	ABL32158	Abl32158 Human imm
	24	65	5.0	7990	6	ABL54307	Abl54307 Chemicall
c	25	64.6	5.0	5286	13	ADS89278	Ads89278 Oligonucl
c	26	64.6	5.0	5286	13	ADS89552	Ads89552 Oligonucl
	27	64.6	5.0	5739	6	ABL32718	Abl32718 Human imm
	28	64.6	5.0	83391	6	ABQ67094	Abq67094 Human ang
	29	64.4	5.0	5286	13	ADS89278	Ads89278 Oligonucl
	30	64.4	5.0	5286	13	ADS89552	Ads89552 Oligonucl
	31	64	4.9	6175	6	ABL33307	Abl33307 Human imm
	32	63.6	4.9	7025	6	ABK40059	Abk40059 Human che
	33	63.6	4.9	7025	6	AAS63350	Aas63350 Chemicall
	34	63.6	4.9	47108	6	ABK31511	Abk31511 Signal tr
c	35	63.4	4.9	2000	11	ACL35887	Acl35887 Rice stre
	36	63.4	4.9	198522	11	ACN44010	Acn44010 Human gen
	37	62.8	4.9	7143	6	ABL32983	Abl32983 Human imm
	38	62.8	4.9	16033	6	ABL33404	Abl33404 Human imm
	39	62.6	4.8	6324	6	ABK31543	Abk31543 Signal tr
	40	62.4	4.8	5689	4	AAS45384	Aas45384 Chemicall
	41	62.4	4.8	5689	4	AAS46426	Aas46426 Tumour su
	42	62.4	4.8	5689	6	ABK28226	Abk28226 DNA trans
	43	62.4	4.8	6775	6	ABQ67159	Abq67159 Human ang
	44	62.2	4.8	1162	14	ADZ71022	Adz71022 Human chr
	45	62.2	4.8	5276	6	ABL32150	Abl32150 Human imm

ALIGNMENTS

RESULT 1
AAX06751
ID AAX06751 standard; DNA; 22080 BP.
XX
AC AAX06751;
XX
DT 26-APR-1999 (first entry)
XX
DE Salmonella enterica serovar B O antigen gene cluster.
XX
KW O antigen; diagnosis; wzx gene; wbaV gene; transferase; polymerase;
KW flippase; diarrhoea; haemorrhagic colitis; ds.
XX
OS Salmonella enterica.
XX
FH Key Location/Qualifiers
FT CDS 4099. .5184
FT /*tag= a
FT /product= "TDP-rhamnose_pathway_protein"
FT /note= "rmlB gene; encodes AAW88335"
FT CDS 5184. .6083
FT /*tag= b
FT /product= "TDP-rhamnose_pathway_protein"
FT /note= "rmlD gene; encodes AAW88336"
FT CDS 6131. .7009
FT /*tag= c
FT /product= "TDP-rhamnose_pathway_protein"
FT /note= "rmlA gene; encodes AAW88337"
FT CDS 7010. .7561
FT /*tag= d
FT /product= "TDP-rhamnose_pathway_protein"
FT /note= "rmlC gene; encodes AAW88338"
FT CDS 7567. .8559
FT /*tag= e
FT /product= "CDP-adequose_pathway_protein"
FT /note= "ddhD gene; encodes AAW88339"
FT CDS 8556. .9329
FT /*tag= f
FT /product= "CDP-adequose_pathway_protein"
FT /note= "ddhA gene; encodes AAW88340"
FT CDS 9334. .10413
FT /*tag= g

```
FT                   /product= "CDP-adequose_pathway_protein"
FT                   /note= "ddhB gene; encodes AAW88341"
FT   CDS             10440. .11753
FT                   /*tag= h
FT                   /product= "CDP-adequose_pathway_protein"
FT                   /note= "ddhC gene; encodes AAW88342"
FT   CDS             11781. .12680
FT                   /*tag= i
FT                   /product= "CDP-adequose_pathway_protein"
FT                   /note= "abe gene; encodes AAW88343"
FT   CDS             12762. .14054
FT                   /*tag= j
FT                   /product= "flippase"
FT                   /note= "wzx gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88344"
FT   CDS             14059. .15060
FT                   /*tag= k
FT                   /product= "abequosyl_transferase"
FT                   /note= "wbaV gene, this region is specifically claimed in
FT                   Claim 9; ; encodes AAW88345"
FT   CDS             15379. .16440
FT                   /*tag= l
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaU gene; encodes AAW88346"
FT   CDS             16441. .17385
FT                   /*tag= m
FT                   /product= "rhamnosyl_transferase"
FT                   /note= "wbaN gene; encodes AAW88347"
FT   CDS             17386. .18825
FT                   /*tag= n
FT                   /product= "GDP_mannose_pathway_protein"
FT                   /note= "manC gene; encodes AAW88348"
FT   CDS             18812. .20245
FT                   /*tag= o
FT                   /product= "GDP_mannose_pathway_protein"
FT                   /note= "mabB gene; encodes AAW88349"
FT   CDS             20317. .21747
FT                   /*tag= h
FT                   /product= "galactosyl_transferase"
FT                   /note= "wbaP gene; encodes AAW88350"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88335, AAW88336, AAW88337, AAW88338, AAW88339, AAW88340,
DR   AAW88341, AAW88342, AAW88343, AAW88344, AAW88345, AAW88346, AAW88347,
DR   AAW88348, AAW88349, AAW88350.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 9; Fig 10; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Salmonella enterica serotype B (strain LT2) O
CC   antigen. 16 Open reading frames within the gene cluster were identified
CC   on the basis of homology to known sequences. The genes encode proteins
CC   (see AAW88335-50) involved the synthesis of sugars present in the
CC   polysaccharide antigen, and in the transport or processing of
CC   polysaccharide or oligosaccharide units. The use of nucleic acid
CC   molecules derived from assembly and transport genes, particularly wbd
CC   (transferase), wzx (flippase) and wzy (polymerase) genes, within O
CC   antigen gene clusters improves the specificity of methods for the
CC   detection and identification of O antigens, e.g. in testing food- or
CC   faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of S. enterica
XX
SQ   Sequence 22080 BP; 6284 A; 4028 C; 5152 G; 6616 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1293;  DB 2;  Length 22080;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-232;
  Matches 1293;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12762 GTGAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTA 12821

Qy         61 TCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12822 TCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTA 12881

Qy        121 GGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12882 GGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCT 12941

Qy        181 GTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12942 GTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAAC 13001

Qy        241 AAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13002 AAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTT 13061

Qy        301 TTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13062 TTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTT 13121

Qy        361 CATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13122 CATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGT 13181

Qy        421 TCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13182 TCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAA 13241

Qy        481 GCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13242 GCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATAC 13301

Qy        541 TATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13302 TATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGT 13361

Qy        601 ATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13362 ATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACA 13421

Qy        661 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13422 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 13481

Qy        721 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13482 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 13541

Qy        781 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13542 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 13601

Qy        841 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13602 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 13661

Qy        901 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 960
```

```
Db      13662 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 13721

Qy        961 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13722 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 13781

Qy       1021 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13782 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 13841

Qy       1081 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13842 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 13901

Qy       1141 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13902 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 13961

Qy       1201 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13962 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCA 14021

Qy       1261 CTAACTTACTTAATTAAGGCAAATAAGGGATAA 1293
              |||||||||||||||||||||||||||||||
Db      14022 CTAACTTACTTAATTAAGGCAAATAAGGGATAA 14054
```

RESULT 2
AEC31432
ID AEC31432 standard; DNA; 14309 BP.
XX
AC AEC31432;
XX
DT 20-OCT-2005 (first entry)
XX
DE E. coli type O43 O-antigen gene cluster.
XX
KW O-antigen; gene expression; DNA detection; gene; ds.
XX
OS Escherichia coli.
XX

FH	Key	Location/Qualifiers
FT	CDS	1083. .2396
FT		/*tag= a
FT		/note= "Orf1"
FT	CDS	2412. .3155
FT		/*tag= b
FT		/note= "Orf2. Specifically claimed in Claim 3"
FT	CDS	3152. .3673
FT		/*tag= c
FT		/note= "Orf3. Specifically claimed in Claim 3"
FT	CDS	3652. .4740
FT		/*tag= d
FT		/note= "Orf4. Specifically claimed in Claim 3"
FT	CDS	4712. .5938
FT		/*tag= e
FT		/note= "Orf5"
FT	CDS	5935. .6984
FT		/*tag= f
FT		/note= "Orf6. Specifically claimed in Claim 3"
FT	CDS	6981. .8099
FT		/*tag= g
FT		/note= "Orf7"
FT	CDS	8103. .9068
FT		/*tag= h
FT		/note= "Orf8"
FT	CDS	9071. .9526
FT		/*tag= i
FT		/note= "Orf9"
FT	CDS	9538. .10941
FT		/*tag= j
FT		/note= "Orf10"
FT	CDS	11007. .11753
FT		/*tag= k
FT		/note= "Orf11. Specifically claimed in Claim 3"
FT	CDS	11778. .13148
FT		/*tag= l
FT		/note= "Orf12"

XX
PN CN1563043-A.
XX
PD 12-JAN-2005.
XX
PF 19-APR-2004; 2004CN-00019028.
XX
PR 19-APR-2004; 2004CN-00019028.
XX
PA (TIAN-) TIANJIN BIOCHIP TECH CO LTD.
XX
PI Wang L, Yang J, Feng L;
XX
DR WPI; 2005-307118/32.
XX
PT Nucleotide peculiar to O-antigen of Escherichia coli O43.
XX
PS Claim 1; SEQ ID NO 1; 35pp; Chinese.
XX
CC The invention relates to a nucleotide sequence (SEQ ID No:1) which is
CC specific to the O-antigen of Escherichia coli O43. The nucleotide
CC sequence represents the complete sequence (14309bp) of a gene cluster for
CC controlling the synthesis of O-antigen in Escherichia coli O43. The
CC nucleotide sequence can have one or several inserted, deleted or
CC substituted bases, whilst still retaining its function. Also described is
CC a PCR primer for the glycosyltransferase (galF) gene and a PCR primer for
CC the oligosaccharide unit treatment (gnd) gene, both of which are
CC contained within the E. coli type O43 O-antigen gene cluster. The
CC invention also discloses a method for detecting and identifying E. coli
CC type O43 in humans. This sequence represents the E. coli type O43 O-
CC antigen gene cluster of the invention.
XX
SQ Sequence 14309 BP; 4529 A; 2267 C; 2972 G; 4541 T; 0 U; 0 Other;

Query Match 20.0%; Score 258.2; DB 14; Length 14309;
Best Local Similarity 51.3%; Pred. No. 5.7e-39;
Matches 639; Conservative 0; Mismatches 588; Indels 18; Gaps 1;

```
Qy         18 AAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTATCCAAAATAATAATTGC 77
              || |||||| |  ||||||   || || | |||||| |||  | | | |   || |  ||
Db       1097 AATAATTCCAAAACATTTATGGGTCGCCGCTTCATCTTGGGGAGCTAGAGGGATTAGCGC 1156

Qy         78 CGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGC 137
                | || ||  | | |||| |    ||| |  | ||| | || || || || |||||| |
Db       1157 TTGTGTACAAATTGTAAGTGTGCATTATTTAGTCTCTTTACTCGGAGAGGATAAATATTC 1216

Qy        138 AATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGG 197
                |||| |||  |    || ||| |   |||||||        ||| || || || ||
Db       1217 TGCTTTTATTTTACTGGGGGGGTTAATTACATGGTGTAATTTATCTGACTTCGGTATCGG 1276

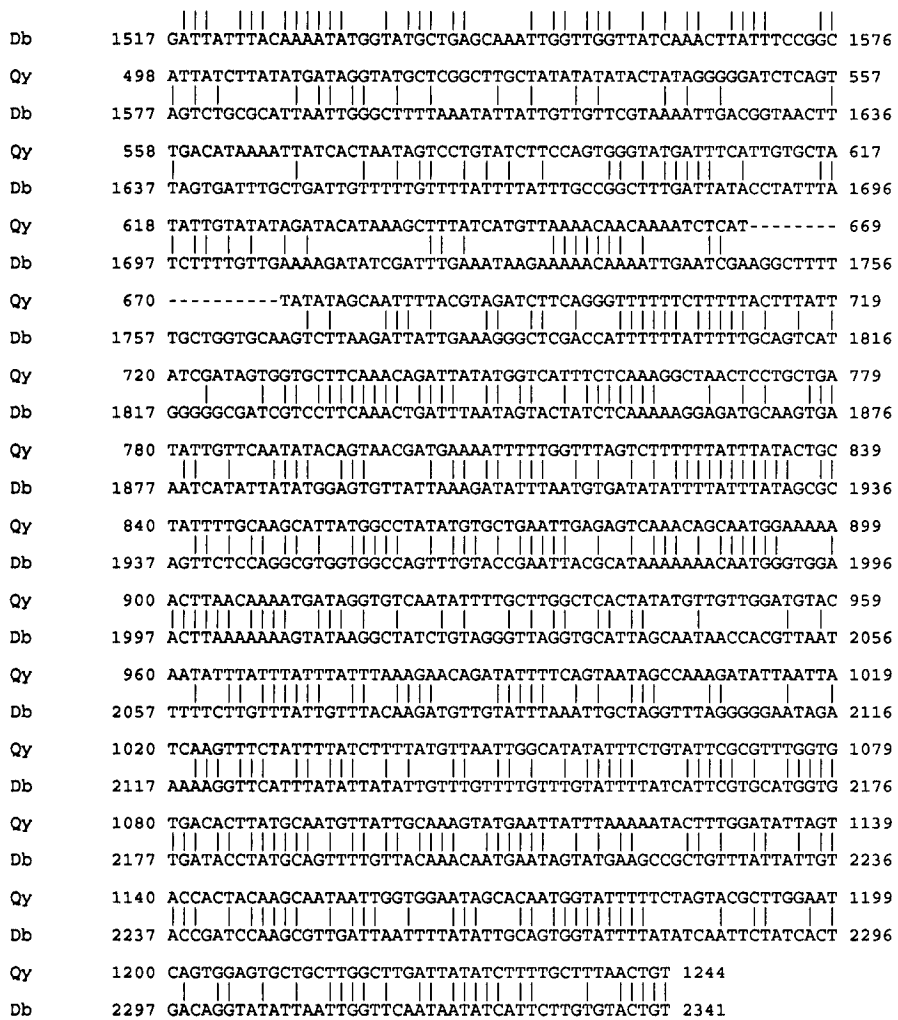
Qy        198 TACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATA 257
                |   |  |  |||||||||  ||||| | || ||  || |||||| ||| ||    |
Db       1277 AAATAGTATATTAAATTATATCGCAGAAAGAAGGGCACAAGACAAAAATTACGAATATTT 1336

Qy        258 TATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTA 317
              |||   |  || ||    | |   ||||||    || || || |    |  ||  |||
Db       1337 TATATTAGTAGGATCTGTTATCGTTTTTATATTATTTTTTATTGTAGCAGTGTTTCTTTG 1396

Qy        318 TATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGA 377
              | |||| ||   |  | |  | |   |||||||       |     || |   |     |
Db       1397 TTTTTTATCAATGACAGTGGCAGAACAATATCTAAAAATATACTCAGATGCTATTATTAA 1456

Qy        378 CAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGC 437
               || |  |  ||  | |||| ||  |||    || ||||  ||  || ||   | |  |
Db       1457 TAATAAAATTATTTTGTTTTATATATCAACATTGATTTTTTGTGGTACTGCGCTTGCCGG 1516

Qy        438 TATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGC 497
```



```
Db       1517 GATTATTTACAAAATATGGTATGCTGAGCAAATTGGTTGGTTATCAAACTTATTTCCGGC 1576

Qy        498 ATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGT 557

Db       1577 AGTCTGCGCATTAATTGGGCTTTTAAATATTATTGTTGTTCGTAAAATTGACGGTAACTT 1636

Qy        558 TGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTA 617

Db       1637 TAGTGATTTGCTGATTGTTTTTGTTTTATTTTATTTGCCGGCTTTGATTATACCTATTTA 1696

Qy        618 TATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCAT-------- 669

Db       1697 TCTTTTGTTGAAAAGATATCGATTTGAAATAAGAAAAACAAAATTGAATCGAAGGCTTTT 1756

Qy        670 ----------TATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATT 719

Db       1757 TGCTGGTGCAAGTCTTAAGATTATTGAAAGGGCTCGACCATTTTTTATTTTTGCAGTCAT 1816

Qy        720 ATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGA 779

Db       1817 GGGGGCGATCGTCCTTCAAACTGATTTAATAGTACTATCTCAAAAAGGAGATGCAAGTGA 1876

Qy        780 TATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGC 839

Db       1877 AATCATATTATATGGAGTGTTATTAAAGATATTTAATGTGATATATTTTATTTATAGCGC 1936

Qy        840 TATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAA 899

Db       1937 AGTTCTCCAGGCGTGGTGGCCAGTTTGTACCGAATTACGCATAAAAAAACAATGGGTGGA 1996

Qy        900 ACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTAC 959

Db       1997 ACTTAAAAAAAGTATAAGGCTATCTGTAGGGTTAGGTGCATTAGCAATAACCACGTTAAT 2056

Qy        960 AATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTA 1019

Db       2057 TTTTCTTGTTTATTGTTTACAAGATGTTGTATTTAAATTGCTAGGTTTAGGGGGAATAGA 2116

Qy       1020 TCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTG 1079

Db       2117 AAAAGGTTCATTTATATTATATTGTTTGTTTTGTTTGTATTTTATCATTCGTGCATGGTG 2176

Qy       1080 TGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGT 1139

Db       2177 TGATACCTATGCAGTTTTGTTACAAACAATGAATAGTATGAAGCCGCTGTTTATTATTGT 2236

Qy       1140 ACCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAAT 1199

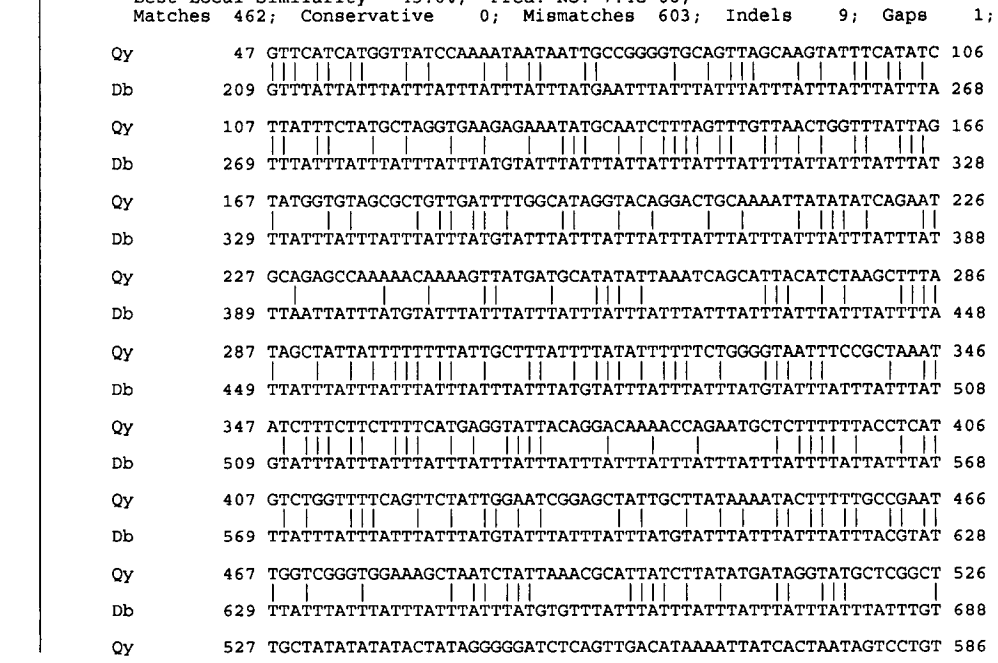
Db       2237 ACCGATCCAAGCGTTGATTAATTTTATATTGCAGTGGTATTTTATATCAATTCTATCACT 2296

Qy       1200 CAGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGT 1244

Db       2297 GACAGGTATATTAATTGGTTCAATAATATCATTCTTGTGTACTGT 2341
```

RESULT 3
ABL33013
ID ABL33013 standard; DNA; 7597 BP.
XX
AC ABL33013;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 986.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 986; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 7597 BP; 1965 A; 109 C; 1790 G; 3733 T; 0 U; 0 Other;

Query Match 6.3%; Score 81.2; DB 6; Length 7597;
Best Local Similarity 43.0%; Pred. No. 7.4e-06;
Matches 462; Conservative 0; Mismatches 603; Indels 9; Gaps 1;



```
Qy         47 GTTCATCATGGTTATCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATC 106

Db        209 GTTTATTATTTATTTATTTATTTATTTATGAATTTATTTATTTATTTATTTATTTATTTA 268

Qy        107 TTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAG 166

Db        269 TTTATTTATTTATTTATTTATGTATTTATTTATTATTTATTTATTTTATTATTTATTTAT 328

Qy        167 TATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAAT 226

Db        329 TTATTTATTTATTTATTTATGTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 388

Qy        227 GCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTA 286

Db        389 TTAATTATTTATGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTTA 448

Qy        287 TAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAAT 346

Db        449 TTATTTATTTATTTATTTATTTATTTATGTATTTATTTATTTATGTATTTATTTATTTAT 508

Qy        347 ATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCAT 406

Db        509 GTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTTATTATTTATTTAT 568

Qy        407 GTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAAT 466

Db        569 TTATTTATTTATTTATTTATGTATTTATTTATTTATGTATTTATTTATTTATTTACGTAT 628

Qy        467 TGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCT 526

Db        629 TTATTTATTTATTTATTTATTTATGTGTTTATTTATTTATTTATTTATTTATTTATTTGT 688

Qy        527 TGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGT 586
```

```
Db        689 TTTATTATTTATTTATTTACGTATTTATATATTTATTTATTTATTTTATTATTCGTTTAT 748

Qy        587 ATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATC 646

Db        749 TTATTTACGTATTTATTTATTTATTTATTTATTTTTTATTTATTTATTTATTTATTTATT 808

Qy        647 ATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTC 706

Db        809 TATTTATTTATTTATTTATTAATTTATGTATTAATTTATGTATTTATTTTATTATTTATT 868

Qy        707 TTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGC 766

Db        869 TATTTATTCGTTTATTCGTTTATTTATTTATTTATTTATGTATTAATTTATTTATTTATT 928

Qy        767 TAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTT 826

Db        929 TATTTATTTATTTTATTATTTATTTATTTATTTATTTATGTATTTATTAGTTTATT 988

Qy        827 TTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAAC 886

Db        989 TATTTATTTATTTATTTATGTATTTATTTATGTATTTATGTATTTATTAATTTATTATTT 1048

Qy        887 AGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATG 946

Db       1049 ATTTATTTAATTATTTATTTATTTATTTAGTTATTAATTTATTTATTTATTTATTTACGT 1108

Qy        947 TTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCA 1006

Db       1109 ATTTATTTATTTATATATTTATTTATTTATTT---------ATTTATTTATTTATCGATA 1159

Qy       1007 AAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTA 1066

Db       1160 TATTTATTTAGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATATATATATA 1219

Qy       1067 TTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAA 1120

Db       1220 TATGTATTTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTGTTTA 1273


RESULT 4
ADL17884/c
ID   ADL17884 standard; cDNA; 158001 BP.
XX
AC   ADL17884;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Human phosphotyrosyl phosphatase activator, PTPA, gene.
XX
KW   Human; ds; antisense; phosphotyrosyl phosphatase activator; PTPA;
KW   hyperproliferative disorder; developmental disorder; infection;
KW   inflammation; tumour; gene.
XX
OS   Homo sapiens.
XX
PN   US2004023906-A1.
XX
PD   05-FEB-2004.
XX
PF   01-AUG-2002; 2002US-00211179.
XX
PR   01-AUG-2002; 2002US-00211179.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Dean NM,  Dobie KW;
XX
DR   WPI; 2004-132607/13.
XX
PT   New antisense compound targeted to a nucleic acid molecule encoding
PT   phosphotyrosyl phosphatase activator, for modulating expression of
PT   phosphotyrosyl phosphatase activator or treating hyperproliferative
PT   disorders.
XX
PS   Example 15; SEQ ID NO 11; 131pp; English.
XX
CC   The invention relates to a compound 8-80 nucleobases in length targeted
CC   to a nucleic acid molecule encoding phosphotyrosyl phosphatase activator
CC   (PTPA), that specifically hybridises with the nucleic acid molecule
CC   encoding phosphotyrosyl phosphatase activator and inhibits the expression
CC   of phosphotyrosyl phosphatase activator, i.e. ana antisense
CC   oligonucleotide. Also included are a composition comprising the compound
CC   and a pharmaceutical carrier or diluent, a method of inhibiting the
CC   expression of phosphotyrosyl phosphatase activator in cells or tissues, a
CC   method of treating an animal having a disease or condition associated
CC   with phosphotyrosyl phosphatase activator and a method of screening for
CC   an antisense compound. The disease or condition is a hyperproliferative
CC   disorder or developmental disorder. The compound, particularly the
CC   antisense oligonucleotide is useful in modulating the function of nucleic
CC   acid molecules encoding phosphotyrosyl phosphatase activator. The
CC   antisense compound can also be used as research tools and diagnostics. It
CC   can also be used as tools in differential and/or combinatorial analyses
CC   to elucidate expression patterns of a portion or the entire complement of
CC   genes expressed within cells and tissues. The compound can also be used
CC   for treating diseases or conditions associated with phosphotyrosyl
CC   phosphatase activator, preferably hyperproliferative disorder or
CC   developmental disorder. The compound can also be used as prophylaxis,
CC   e.g. to prevent or delay infection, inflammation or tumour formation. The
CC   present sequence is the human PTPA gene which is a target for the
CC   antisense oligonucleotides of the invention.
XX
SQ   Sequence 158001 BP; 34764 A; 40497 C; 41696 G; 40433 T; 0 U; 611 Other;

  Query Match             5.8%;  Score 74.8;  DB 12;  Length 158001;
  Best Local Similarity   44.7%;  Pred. No. 0.00014;
  Matches  388;  Conservative    0;  Mismatches  467;  Indels   13;  Gaps    2;

Qy        255 ATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATT 314

Db      23024 ATATATTATATATTATATATTATATATTTTTATATATTATATATTATATATTTATATTT 22965

Qy        315 TTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACA 374

Db      22964 TTATATATTAT-----ATATTATATATTATATATTTTTATATATTATATATTATATATTT 22910

Qy        375 GGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGG 434

Db      22909 ATATATTATATATTGTATATATTTTTATATATTATATATTATATATTATATATTTTTATA 22850

Qy        435 AGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAA 494

Db      22849 TATTATATATTATATATTTATATATTATATATTGTATATTTTTTTATATATTATATATTA 22790

Qy        495 CGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTC 554

Db      22789 TATATTTTTATATATTATATATTATATATTTATATATTATATATTGTATATTTTTTTATA 22730

Qy        555 AGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTG 614

Db      22729 TATTATATATTATATATTATATATTTTTATATATTATATATTATATATTTATATATTATA 22670

Qy        615 CTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATAT 674

Db      22669 TATTGTATATTTTTTTATATATTATATATTTTATATATTTATATATAATATATATTATAT 22610

Qy        675 AGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCT 734

Db      22609 ATTTTATATATTTATATATAATATATTATATATTTTATATATTTATATATAATATATA 22550

Qy        735 TCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATAC 794

Db      22549 TTATATATTTTATATATTTATATATAATATATATTATATATTTTATATATTTATATATAA 22490

Qy        795 AGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATT 854
```

```
Db      22489 TATATATTATATATTTTATATATTTATATATAATATATATTATATATTTTATATATTTAT 22430

Qy        855 ATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGAT 914
              ||    ||||| |  |  |||      |          |     | | |    ||         ||
Db      22429 ATATAATATATATTATATATTTTATATATTTATATATAATATATATTATATATTTTATAT 22370

Qy        915 AGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTT 974
              |   | |  ||| | |    |      |   | | |    |  || || ||||| ||||   |
Db      22369 ATTTATATATAATATATTATATATATTTTATATATTTATATATAATATATTATTATA 22310

Qy        975 ATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTT 1034
              ||||  | |        |||   ||  |   | |      || ||    |  ||| ||| ||
Db      22309 TTTTATATA--------TTTATATATAATATATATTATATATTTTATATATTTATATATT 22258

Qy       1035 ATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAAT 1094
              || |||||| |||  |   |||||| |    |||      |    | | |  |||   ||
Db      22257 ATTTTTTATATTATATATTATATATATTACAATTATATAATATATTAATATATATTATAT 22198

Qy       1095 GTTATTGCAAAGTATGAATTATTTAAAA 1122
              |||||  | | |||  ||||| || ||
Db      22197 ATTATTATATATTATTTATTATATATAA 22170


RESULT 5
ADZ71010
ID   ADZ71010 standard; DNA; 842 BP.
XX
AC   ADZ71010;
XX
DT   14-JUL-2005  (first entry)
XX
DE   Human chromosome 2 contig DNA SEQ ID NO 134.
XX
KW   ds; matrix attachment region; MAR; protein production.
XX
OS   Homo sapiens.
XX
PN   WO2005040377-A2.
XX
PD   06-MAY-2005.
XX
PF   22-OCT-2004; 2004WO-EP011974.
XX
PR   24-OCT-2003; 2003US-0513574P.
PR   06-FEB-2004; 2004EP-00002722.
XX
PA   (SELE-) SELEXIS SA.
XX
PI   Mermod N,  Girod PA,  Bucher P,  Nguyen D,  Calabrese D,  Saugy D;
PI   Puttini S;
XX
DR   WPI; 2005-333507/34.
XX
PT   New purified and isolated DNA sequence having protein production
PT   increasing activity comprises a bent DNA element and a binding site for a
PT   DNA binding protein, useful for increasing protein production activity in
PT   eukaryotic host cell.
XX
PS   Disclosure; SEQ ID NO 134; 282pp; English.
XX
CC   The invention relates to a purified and isolated DNA sequence having
CC   protein production increasing activity comprising at least one bent DNA
CC   element, and at least one binding site for a DNA binding protein. The
CC   purified and isolated DNA sequence comprising a first and a second
CC   isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC   MAR nucleotide sequence selected from a purified and isolated DNA
CC   sequence above, a purified and isolated MAR DNA above, a purified and
CC   isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC   sequence complementary to it, its molecular chimera, or its combinations
CC   and variants, is useful for increasing protein production activity in a
CC   eukaryotic host cell. The present sequence represents a human chromosome
CC   2 contig DNA.
XX

SQ   Sequence 842 BP; 351 A; 7 C; 20 G; 464 T; 0 U; 0 Other;

  Query Match             5.7%;  Score 73.8;  DB 14;  Length 842;
  Best Local Similarity   46.3%;  Pred. No. 0.00015;
  Matches  281;  Conservative    0;  Mismatches  322;  Indels    4;  Gaps    1;

Qy        534 TATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCC 593
              |||||| |    |    || |   || | | | || |||    |    ||    ||| |
Db          1 TATATATTTATGTTTAATATATATTTATAGATAAATATATATTTACGTTTAATATATATT 60

Qy        594 AGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAA 653
              | | ||    |  |||    | || | |||||  ||  |||||  | | |    ||||
Db         61 TATAGATAAATATATATTTACGTTTAATATATATTTATCTATAAATATATTTACGTTTAA 120

Qy        654 AACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTAC 713
              | |    || |  | |||||   || ||   || || || |   | || | | |
Db        121 TATATATTTATATATTAATATATTTATGTTTAATATATATTTATATATATTAATATATTT 180

Qy        714 TTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCC 773
              |  |||        |    |   || | |||      | || | |  |    ||
Db        181 ATGTTTAATATATATTTATATATTAATATATTTATGTTTAATATATTTATATATATTAAT 240

Qy        774 TGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTA 833
                |     | |   |||||     | | | | |  ||| ||| ||    | ||||| |
Db        241 ATATTTATGTTTAATATATATTTATATGTTAATATATTTAGGTATATATATATTTATATG 300

Qy        834 TACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATG 893
              |    ||| ||   |  | |||    ||| | |    ||  |   |  || |   |
Db        301 TTAATATATATTTATATTAATATATTATATTTATATATAAAAGTATATATAATATATAAA 360

Qy        894 GA----AAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTG 949
              |    | ||| |     | |  ||    | |  ||||||    |      ||||| ||
Db        361 TATTATATAAATTATTATATAGTATTTTTATATATATTTATATATAAATTTTATATATTT 420

Qy        950 TTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAG 1009
              |          ||||| ||||||| |||  |    | | |||  |  |    ||   | |
Db        421 TATATATATAAATATATATTTATATATACATTTTATATATAAATATATATTTATATATAC 480

Qy       1010 ATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTC 1069
              ||  ||  ||| ||  |     ||||| |||||  |  |      |||||||  | ||| |
Db        481 ATTATATATATAAATATATATTTATATTTTATATATAAATATATATATTTATATATAC 540

Qy       1070 GCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTT 1129
                ||   | |   |  |||| ||   |||    || | | |  ||||  | |    |||
Db        541 ATTTTATATATTTTATATATGTAAATATATATATAAATTTTATATATTGTATATATATTT 600

Qy       1130 GGATATT 1136
                | |||
Db        601 ATAAATT 607


RESULT 6
ADZ70897/c
ID   ADZ70897 standard; DNA; 1524 BP.
XX
AC   ADZ70897;
XX
DT   14-JUL-2005  (first entry)
XX
DE   Human matrix attachment region DNA SEQ ID NO 21.
XX
KW   ds; matrix attachment region; MAR; protein production.
XX
OS   Homo sapiens.
XX
PN   WO2005040377-A2.
XX
PD   06-MAY-2005.
XX
PF   22-OCT-2004; 2004WO-EP011974.
XX
PR   24-OCT-2003; 2003US-0513574P.
PR   06-FEB-2004; 2004EP-00002722.
```

XX
PA (SELE-) SELEXIS SA.
XX
PI Mermod N, Girod PA, Bucher P, Nguyen D, Calabrese D, Saugy D;
PI Puttini S;
XX
DR WPI; 2005-333507/34.
XX
PT New purified and isolated DNA sequence having protein production
PT increasing activity comprises a bent DNA element and a binding site for a
PT DNA binding protein, useful for increasing protein production activity in
PT eukaryotic host cell.
XX
PS Claim 4; SEQ ID NO 21; 282pp; English.
XX
CC The invention relates to a purified and isolated DNA sequence having
CC protein production increasing activity comprising at least one bent DNA
CC element, and at least one binding site for a DNA binding protein. The
CC purified and isolated DNA sequence comprising a first and a second
CC isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC MAR nucleotide sequence selected from a purified and isolated DNA
CC sequence above, a purified and isolated MAR DNA above, a purified and
CC isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC sequence complementary to it, its molecular chimera, or its combinations
CC and variants, is useful for increasing protein production activity in a
CC eukaryotic host cell. The present sequence represents a human matrix
CC attachment region DNA.
XX
SQ Sequence 1524 BP; 943 A; 14 C; 15 G; 552 T; 0 U; 0 Other;

Query Match 5.6%; Score 72.4; DB 14; Length 1524;
Best Local Similarity 44.4%; Pred. No. 0.00029;
Matches 461; Conservative 0; Mismatches 571; Indels 6; Gaps 4;

```
Qy         96 TATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTTAAC 155
Db       1414 TATATCTTTATATATCTTTATATATGTTTATTTATATATATTTATATATGTTTATATATT 1355

Qy        156 TGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAATTA 215
Db       1354 TATATTTATATATTTATATATATTTATATATTTATATTTATATATTTATATATATTTATA 1295

Qy        216 TATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTACA 275
Db       1294 TATATTTATATATTTTTATATATTTATATATTTATATATATTTATACATTTATATATTTA 1235

Qy        276 TCTA-AGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAA 334
Db       1234 TATATATTTATATATATATTTATATATTTATATATATTTTTTATATTTATATTTATATAT 1175

Qy        335 TTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCT 394
Db       1174 TTATATATATATTTATATTTATATATTGTTATATATTTTTATATATATTTATATTTATAT 1115

Qy        395 TTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATAC 454
Db       1114 ATTGTTATATATTTTTATATATTTA-TATATTGTTATATATTTTATATATTTATATATTGT 1056

Qy        455 TTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAG 514
Db       1055 TATATATTTTTATATATATTTATACATTTATATATTTATATATATTTATATATTTATATA 996

Qy        515 GTATGCTCGGCTTGCTATATATATA---CTATAGGGGGATCTCAGTTGACATAAAATTAT 571
Db        995 TATTTTTTATATTTATATTTATATATTTATATATATTTATATTTATATATTGTTATATAT 936

Qy        572 CACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGAT 631
Db        935 TTTTATATATATTTATATTTATATATTGTTATATATTTTTTATATATATTTATATATTGTT 876

Qy        632 ACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGAT 691
Db        875 ATATATTTTTATATATTTATATATTTATATATTTATATATATTTATATATATTTATTAT 816
```

```
Qy        692 CTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGG 751
Db        815 ATATTTATATATTTATATATATTTATATATATTTATATATATTTATATATATTTATATAT 756

Qy        752 TCATTTCTCAAAGGCTA-ACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATT 810
Db        755 TTATATATATTTATATATATTTATATATTTATATTTATATATTTATATATTTATATATAT 696

Qy        811 TTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCT 870
Db        695 TTATATATATTTATATATTTATATATTTATATATATTTATATATATTTATATATTTATAT 636

Qy        871 GAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTG 930
Db        635 ATATTTATATATATTTATATATTTATATATATTTATATATTTATATATATTTTTTATATAT 576

Qy        931 CTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATA 990
Db        575 TTTTATATATTTATATTTATATATATTTATATATTTATATATATTTTTTATATATTTATA 516

Qy        991 TTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATT 1050
Db        515 TATTCATATATATTCATAAATACTTATATTTATTTATTTATATATGTATATATGTATATA 456

Qy       1051 GGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATG 1110
Db        455 TTTATATATTTATATATATTTATATATTTATATATATTTTTTATATATTTATATATTCAT 396

Qy       1111 AATTATTTAAAAATACTT 1128
Db        395 ATATATTCATAAATACTT 378
```

RESULT 7
ADZ70902/c
ID ADZ70902 standard; DNA; 4660 BP.
XX
AC ADZ70902;
XX
DT 14-JUL-2005 (first entry)
XX
DE Human matrix attachment region DNA SEQ ID NO 26.
XX
KW ds; matrix attachment region; MAR; protein production.
XX
OS Homo sapiens.
XX
PN WO2005040377-A2.
XX
PD 06-MAY-2005.
XX
PF 22-OCT-2004; 2004WO-EP011974.
XX
PR 24-OCT-2003; 2003US-0513574P.
PR 06-FEB-2004; 2004EP-00002722.
XX
PA (SELE-) SELEXIS SA.
XX
PI Mermod N, Girod PA, Bucher P, Nguyen D, Calabrese D, Saugy D;
PI Puttini S;
XX
DR WPI; 2005-333507/34.
XX
PT New purified and isolated DNA sequence having protein production
PT increasing activity comprises a bent DNA element and a binding site for a
PT DNA binding protein, useful for increasing protein production activity in
PT eukaryotic host cell.
XX
PS Claim 4; SEQ ID NO 26; 282pp; English.
XX
CC The invention relates to a purified and isolated DNA sequence having
CC protein production increasing activity comprising at least one bent DNA

```
CC   element, and at least one binding site for a DNA binding protein. The
CC   purified and isolated DNA sequence comprising a first and a second
CC   isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC   MAR nucleotide sequence selected from a purified and isolated DNA
CC   sequence above, a purified and isolated MAR DNA above, a purified and
CC   isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC   sequence complementary to it, its molecular chimera, or its combinations
CC   and variants, is useful for increasing protein production activity in a
CC   eukaryotic host cell. The present sequence represents a human matrix
CC   attachment region DNA.
XX
SQ   Sequence 4660 BP; 1770 A; 774 C; 775 G; 1341 T; 0 U; 0 Other;

  Query Match             5.6%;  Score 72.4;  DB 14;  Length 4660;
  Best Local Similarity   44.4%;  Pred. No. 0.00032;
  Matches  461;  Conservative    0;  Mismatches  571;  Indels    6;  Gaps    4;

Qy         96 TATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTTAAC 155
Db       2423 TATATCTTTATATATCTTTATATATGTTTATTTATATATATTTATATATGTTTATATATT 2364

Qy        156 TGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAATTA 215
Db       2363 TATATTTATATATTTATATATATTTATATATTTATATTTATATATTTATATATATTTATA 2304

Qy        216 TATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTACA 275
Db       2303 TATATTTATATATTTTTATATATTTATATATTTATATATATTTATACATTTATATATTTA 2244

Qy        276 TCTA-AGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAA 334
Db       2243 TATATATTTATATATATATTTATATATTTATATATATTTTTTATATTTATATTTATATAT 2184

Qy        335 TTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCT 394
Db       2183 TTATATATATATTTATATTTATATATTGTTATATATTTTTATATATATTTATATTTATAT 2124

Qy        395 TTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATAC 454
Db       2123 ATTGTTATATATTTTTATATATTTA-TATATTGTTATATATTTATATATTTATATATTGT 2065

Qy        455 TTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAG 514
Db       2064 TATATATTTTTATATATATTTATACATTTATATATTTATATATATTTATATATTTATATA 2005

Qy        515 GTATGCTCGGCTTGCTATATATATA---CTATAGGGGGATCTCAGTTGACATAAAATTAT 571
Db       2004 TATTTTTTATATTTATATTTATATATTTATATATATTTATATTTATATATTGTTATATAT 1945

Qy        572 CACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGAT 631
Db       1944 TTTTATATATATTTATATTTATATATTGTTATATATTTTTATATATATTTATATATTGTT 1885

Qy        632 ACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGAT 691
Db       1884 ATATATTTTTATATATTTATATATTTATATATTTATATATATTTATATATATATTTATAT 1825

Qy        692 CTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGG 751
Db       1824 ATATTTATATATTTATATATATTTATATATATTTATATATATTTATATATATTTATATAT 1765

Qy        752 TCATTTCTCAAAGGCTA-ACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATT 810
Db       1764 TTATATATATTTATATATATTTATATATTTATATTTATATATTTATATATTTATATATAT 1705

Qy        811 TTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCT 870
Db       1704 TTATATATATTTATATATTTATATATTTATATATATTTATATATATTTATATATTTATAT 1645

Qy        871 GAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTG 930
Db       1644 ATATTTATATATATTTATATATTTATATATATTTATATATTTATATATATTTTTATATAT 1585

Qy        931 CTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATA 990
Db       1584 TTTTATATATTTATATTTATATATATTTATATATTTATATATATTTTTTATATATTTATA 1525

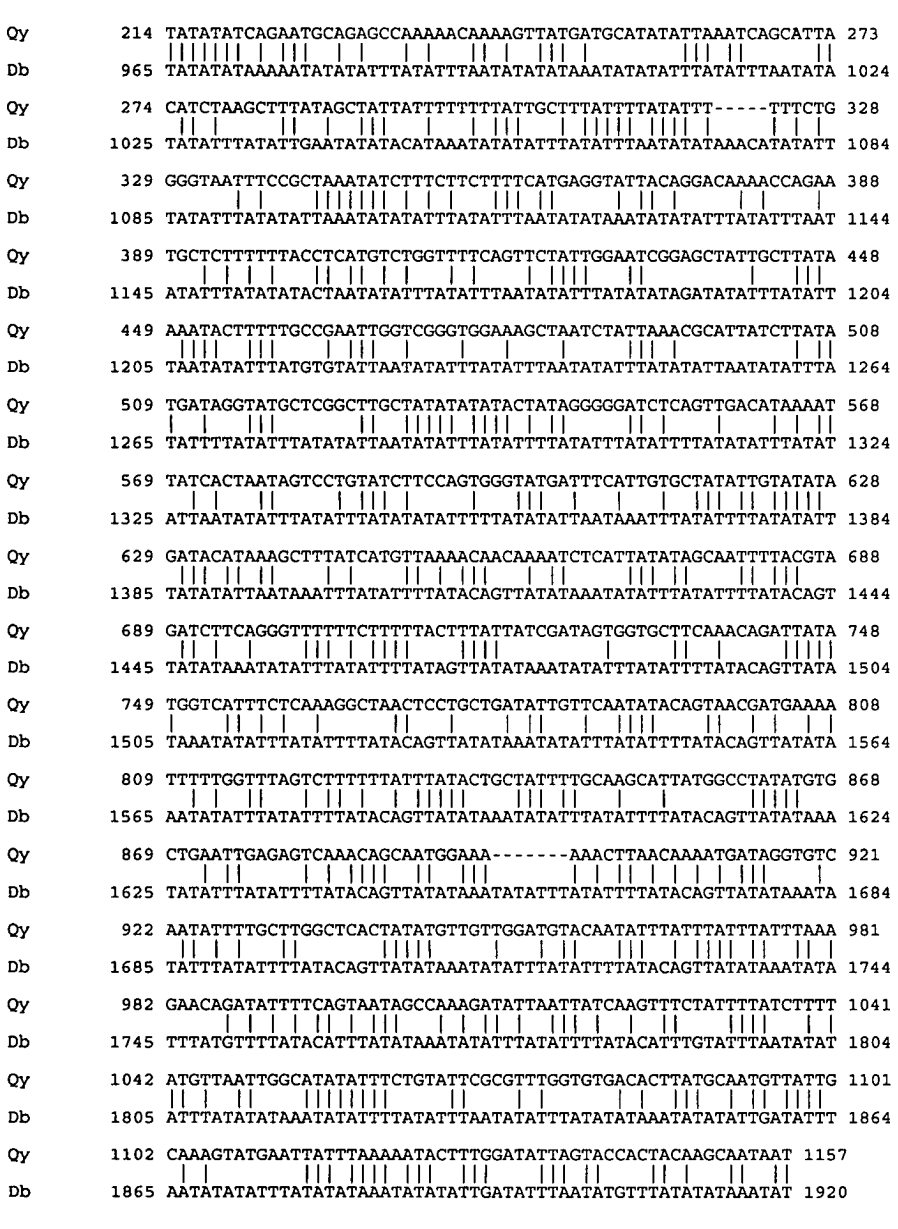
Qy        991 TTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATT 1050
Db       1524 TATTCATATATATTCATAAATACTTATATTTATTTATTTATATATGTATATATGTATATA 1465

Qy       1051 GGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATG 1110
Db       1464 TTTATATATTTATATATATTTATATATTTATATATATTTTTTATATATTTATATATTCAT 1405

Qy       1111 AATTATTTAAAAATACTT 1128
Db       1404 ATATATTCATAAATACTT 1387


RESULT 8
ADZ71009
ID   ADZ71009 standard; DNA; 2131 BP.
XX
AC   ADZ71009;
XX
DT   14-JUL-2005  (first entry)
XX
DE   Human chromosome 2 contig DNA SEQ ID NO 133.
XX
KW   ds; matrix attachment region; MAR; protein production.
XX
OS   Homo sapiens.
XX
PN   WO2005040377-A2.
XX
PD   06-MAY-2005.
XX
PF   22-OCT-2004; 2004WO-EP011974.
XX
PR   24-OCT-2003; 2003US-0513574P.
PR   06-FEB-2004; 2004EP-00002722.
XX
PA   (SELE-) SELEXIS SA.
XX
PI   Mermod N,  Girod PA,  Bucher P,  Nguyen D,  Calabrese D,  Saugy D;
PI   Puttini S;
XX
DR   WPI; 2005-333507/34.
XX
PT   New purified and isolated DNA sequence having protein production
PT   increasing activity comprises a bent DNA element and a binding site for a
PT   DNA binding protein, useful for increasing protein production activity in
PT   eukaryotic host cell.
XX
PS   Disclosure; SEQ ID NO 133; 282pp; English.
XX
CC   The invention relates to a purified and isolated DNA sequence having
CC   protein production increasing activity comprising at least one bent DNA
CC   element, and at least one binding site for a DNA binding protein. The
CC   purified and isolated DNA sequence comprising a first and a second
CC   isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC   MAR nucleotide sequence selected from a purified and isolated DNA
CC   sequence above, a purified and isolated MAR DNA above, a purified and
CC   isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC   sequence complementary to it, its molecular chimera, or its combinations
CC   and variants, is useful for increasing protein production activity in a
CC   eukaryotic host cell. The present sequence represents a human chromosome
CC   2 contig DNA.
XX
SQ   Sequence 2131 BP; 923 A; 38 C; 31 G; 1139 T; 0 U; 0 Other;

  Query Match             5.6%;  Score 72;  DB 14;  Length 2131;
  Best Local Similarity   43.8%;  Pred. No. 0.00035;
  Matches  419;  Conservative    0;  Mismatches  525;  Indels   12;  Gaps    2;
```



RESULT 9
ABL32281
ID ABL32281 standard; DNA; 13511 BP.
XX
AC ABL32281;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 254.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 254; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 13511 BP; 4605 A; 85 C; 2513 G; 6307 T; 0 U; 1 Other;

Query Match 5.6%; Score 71.8; DB 6; Length 13511;
Best Local Similarity 44.5%; Pred. No. 0.00044;
Matches 370; Conservative 0; Mismatches 457; Indels 4; Gaps 2;

```
Qy        657 AACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTT 716
              |    |||| | ||  |||||  ||| ||  |  || ||  |  ||||||  ||||  |
Db      11124 ATAGAAATATTATATTATAGAGATTATATTTTTATTTTTTGATTTTTTTTATTTTTTTAG 11183

Qy        717 ATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGC 776
              ||  |  |||   | | |  ||    ||| | |  | | || | |      || |  |
Db      11184 ATGTTAAATATATGAGGT--AATTTTATTTTTTATTTAGTTTTTATTTTTGAAGTATTTT 11241

Qy        777 TGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATAC 836
              |  |||   |   |     |    |   ||| |   | | ||  |  || | | |
Db      11242 TTTTATGTATTTTTTATTTTTTTATTTTGTTATTATGTTTATGGTTATAGTTTTATTTTT 11301

Qy        837 TGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAA 896
              | |||| || ||  ||  |          |||  |         |||          |
Db      11302 AGTTATTATGAAATTATATTTTGAAGATTGTATTAGTGGTTTAGTTGGTTTTTGAGTTTT 11361

Qy        897 AAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATG 956
              | || ||          |   | | | |  |||  |  |   | | ||   | |    |
Db      11362 AGAATTTTTTTTGTAATTTTTTTTAATTTATTTTGTAAGGATATTTTAGAGTTTATTTTT 11421

Qy        957 TACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAA 1016
              ||  |   | ||||  ||    |    |  | ||||||    |  |    | || ||||
Db      11422 TAGTAGTATAATTTTATTTAAAATTTTAGTGTTATTTTTTTAAGGAAATTACGAAATTAG 11481

Qy       1017 TTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTG 1076
              ||||  |||||   ||||   | ||         |||    || ||  | |||      |||
Db      11482 TTATAGAGTTTTAGTTTTTAATATTGGAAATTTTGAGTAATTTTATTTTATATTATTTTT 11541

Qy       1077 GTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACT 1127
                |    | || |    | |||||    |   |  |||||||   | | |
Db      11542 TAGATTTATTTTTTTTTTTTTATTTTGTATAGTTTATTATTTAGTTAAATT 11592
```

RESULT 10
ABL34155
ID ABL34155 standard; DNA; 15548 BP.
XX
AC ABL34155;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 2128.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 2128; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 15548 BP; 4209 A; 247 C; 2903 G; 8189 T; 0 U; 0 Other;

Query Match 5.5%; Score 71; DB 6; Length 15548;
Best Local Similarity 44.9%; Pred. No. 0.00063;
Matches 470; Conservative 0; Mismatches 570; Indels 7; Gaps 5;

```
Qy         96 TATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTTAAC 155
              ||| |  ||| ||  || |||  |      | |  |||   || ||||  || |
Db      11675 TATATTTTATATTTATTTTATTTTTATATTTTATATTTATTTTATTTTTATATTTTATTT 11734

Qy        156 TGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAATTA 215
              |  | | ||| |||  | |      | ||   |||    |   ||     | |   | |||
Db      11735 TTATATTTTATTATTATATTTTATTTTTATATTTTATTTTTATATTTTATTTTTATATTT 11794

Qy        216 TATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTACA 275
              ||| |   ||         | |    | ||   ||     | ||| ||     ||  |
Db      11795 TATTTTTATATTTTATTTTTATATTTTATTTTTATATTTTATTTTTATATTTTATTTTTA 11854

Qy        276 TCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAAT 335
              | | |  |||||  |  ||||| |||||||  | ||||||||||||| || |    | ||
Db      11855 TTTTATTTTTATATTTTATATTTATTTTATTTTTATATTTTATATTTATTTTATTTTTAT 11914

Qy        336 TTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTT 395
              |    |   ||| |||  | |||  |||   | ||         | |  | | |    ||
Db      11915 ATTTTATTTTTATATTTTATTTTTATATGTTATTTTTTATATGTTATA-TATATTTTATTT 11973

Qy        396 TTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACT 455
              || ||  | || | |  |||    || |||   |      |  |  |  |||||   || |
Db      11974 TTATATGTTATATATATTTTATTTTTATATGTTATATATATTTTATTTTTATATTTTATT 12033

Qy        456 TTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGG 515
              |||    |  |  |   || | | || | |||    ||   |  | | ||| |
Db      12034 TTTATATGTTATTTTTATATGTTATTTTTATATGTTATTTTTATATGTTATTTTTATATG 12093

Qy        516 TATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACT 575
              |    |   || ||| | ||||  ||    ||   |  | | ||| |  | | | |
Db      12094 TTATTTTTATATGTTATTTTTATATGTTATTTTTAT---ATGTTATATATATATGTTATT 12150

Qy        576 AATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACAT 635
                ||    | ||| |    ||        || | |     |||||  |||||| |||  |
Db      12151 TTTATAGTTATATATATATGTTATATATATATGTTATATATATATGTTATATATATATGT 12210

Qy        636 AAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTC 695
              |  |  |  |  || |  | |  | |  |||||  |  | || ||  | |  |
Db      12211 TATATATATATGTTATATATATATATGTTATATATATATGTTATATATATATATGTTATA 12270

Qy        696 AGGGTTTTTTCTTTTTA-CTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCA 754
                | | || | | ||  | | ||||  |||   ||  |  | | | ||  |||    |
Db      12271 TATATATGTTATATATATATATGTTATATATATATGTTATATATATATATGTTATATATA 12330

Qy        755 TTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTG 814
              | | | | |      |   | |      | |  |||||  |   | | |  | || | |
Db      12331 TATGTTATATATATATATGTTATATATATATGTTATATATATATATGTTATATATATATA 12390

Qy        815 GTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAAT 874
              |||  | | | |  |||||   |||  ||  | | ||     |   |    |    |
Db      12391 TGTTA-TGTATATATGTTATATATATATATTTTATGTATATATGTTATATATATATATTT 12449

Qy        875 TGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTG 934
              | |  |  |  |  || | |  |||  | ||  |  | |  |||| ||   ||
Db      12450 TATGTATATATTTTATATATATATATTTTATGTATATATTTTATATATATATATTTTATG 12509
```

```
Qy        935 GCT-CACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTT 993
              |  | | ||| |   |  || |    ||| |||||  |   |  |           |
Db      12510 TATATATTTTATATATATATATTTTATGTATATATTTTATATATATATATTTTATGTATA 12569

Qy        994 TCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGC 1053
              |   | |||   | | || |||  |||  |  || ||| |     ||     ||  |  |
Db      12570 TATTTTATATATATATATTTTATGTATATATGTTATATATATATATTATATATATATTTC 12629

Qy       1054 ATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAAT 1113
              ||||| | | | ||     |   |  | | ||||   | |   |  |   |||  ||
Db      12630 GTATATATATATTTTATATATATATTATATTTATTTCGTATATATATATTTTATATAT 12689

Qy       1114 TATTTAAAAATACTTTGGATATTAGTA 1140
                ||| | |||  ||   |||   ||
Db      12690 ATATTATATATATATTTCGTATATATA 12716
```

RESULT 11
AAS46788
ID AAS46788 standard; DNA; 61020 BP.
XX
AC AAS46788;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #514.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 514; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 61020 BP; 17884 A; 634 C; 12537 G; 29965 T; 0 U; 0 Other;

```
  Query Match             5.5%;  Score 71;  DB 4;  Length 61020;
  Best Local Similarity   44.9%;  Pred. No. 0.00069;
  Matches  484;  Conservative    0;  Mismatches  580;  Indels   15;  Gaps    5;

Qy         34 TTAATTGTTGCAGGTTCATCATGGTTATCCAAAATAATAATTGCCGGGGTGCAGTTAGCA 93
              | | || | | |  ||  | |   |||       ||| || ||            ||
Db      33944 TAAGTTATAGAAATTTTTTTAATTTAGTTTTTATATTAGTTAAATTATATTTTTTTTAT 34003

Qy         94 AGTATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTTA 153
              || |||| || | || | | | |   | |  | ||  |||||  | | |  |  | ||
Db      34004 AGGATTTTATTTTTTTTGTGTTTTTAATGAAATGATGAATATAAATTATGAATATCGTGT 34063

Qy        154 ACTGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAAT 213
              ||  |   | || |||  |  |  || | | ||   ||   ||  |||     |  ||
Db      34064 TTTGTATATATTGTTTGGTTAAAAG--GTAGTTATTATTATTATTGGATATTATTAT 34121

Qy        214 TATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTA 273
              |||   | | | |    |       |    | ||     |     | ||   | |   |||
Db      34122 TATTATTGAAATTAATTAAATTTTGATTTTATTTTGATTAATGGTTTTTTTTTTATTTTTA 34181

Qy        274 CATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTA 333
              || |||   | ||   |  |  || |||||||    ||||||  ||| |||| | |  ||
Db      34182 AATGTAA--ATAATTATTTGTTGTTATTTTATTTTGTTATTTTTTATATTTTATTGTTTA 34239

Qy        334 ATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTC 393
              |   | |||||  ||| || || |  |   || ||   | | | |       |
Db      34240 TTCGTTTTTTAATATTTTTGTTTTTGTTTTTTAGTTTTTTTTGTAAATATTTTTTGTAAGT 34299

Qy        394 TTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATA 453
              |||  |   |  | | |  | || | | || |  | | |    | ||| ||||||  |
Db      34300 TTTAATTTGTTTTATTTTATGGTTATTTTTTTTTAAGTTGTTTTTTTTGTTTATAATGTT 34359

Qy        454 CTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATA 513
              |||        |||  |    |  ||     |      ||          | | || || ||||
Db      34360 TTTTCGATTTTATTATTATTTTTAAATTTAATAAATTT----TTTATTTTTTTTAAGATA 34415

Qy        514 GGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCA 573
              | |       ||  |   | |||| || |    || |    | |  | | ||
Db      34416 TTTTTAAGGATATTTTTGGTTTATAGTAAAATAAAATATAGAGTTATTTTATATATATT 34475

Qy        574 CTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATAC 633
                ||||  | | | |   | |  ||| |||    |||  |||     || || ||
Db      34476 TATTTAGTTATATGTATAGTATTTTTTATTATTAATATGTTTTATTAGAAATGTATATTT 34535

Qy        634 ATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCT 693
              || |   ||||| |  ||| | | |   | | | ||||| ||    || || |      |
Db      34536 ATTATAATTTATGAATTTATATCGATGTATTATTATTATTTAAAGTTTATAGTTTATATT 34595

Qy        694 TCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTC 753
              || | |||| | | |  | |||| |     | |              || |||||
Db      34596 AGGGGTTATTTTTTGTGTTGTATATTTTATGGGTTTGGATAAATGGATAATGATATGTAT 34655

Qy        754 ATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACA-----GTAACGATGAAAA 808
              | |     |  |    |  |  ||||   | | | || |       | | | |  |
Db      34656 TTATTATTATAGTATTATATAGAATATATATTATTTTAAAAATTTTTTATGTTTTATT 34715

Qy        809 TTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTG 868
              | |||  |||  |  ||||  ||| |  || || ||    |   |  | |  | | | |
Db      34716 TATTTATTTTTTTTGTTTTATTTTTTTTTGATAATTATTGATTTTATTAGTATTTTTATA 34775

Qy        869 CTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTT 928
                |    |     | |        |||||  |  | |   | ||  |   |    |||||
```

```
Db      34776 GTTTTGTTTTTTTTAGTATATGTTGGAATTATATAATATATATTTTTTTTAGATTTATTT 34835

Qy        929 TGCTTGGCTCACTATATGTTGTT--GGATGTACAATATTTATTTATTTATTTAAAGAACA 986

Db      34836 TTTTTATTTAGTAATATGTATTTAAGTTTTTTTTATGTTTTTTTATGGTTTGATAGTTTA 34895

Qy        987 GATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTT 1046

Db      34896 TTTTTTTTTAGTTTTGAATAATATTTTTTTGTTTAGATATTATAGTTTATTAGTTTATTT 34955

Qy       1047 AATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAA 1105

Db      34956 ATTGAAGGATATTTTAGTTGTTTTCGAGTTTTGGTAATTATGAATAAAGTTGTTATAAA 35014
```

RESULT 12
ADZ70929/c
ID ADZ70929 standard; DNA; 828 BP.
XX
AC ADZ70929;
XX
DT 14-JUL-2005 (first entry)
XX
DE Human chromosome 1 contig DNA SEQ ID NO 53.
XX
KW ds; matrix attachment region; MAR; protein production.
XX
OS Homo sapiens.
XX
PN WO2005040377-A2.
XX
PD 06-MAY-2005.
XX
PF 22-OCT-2004; 2004WO-EP011974.
XX
PR 24-OCT-2003; 2003US-0513574P.
PR 06-FEB-2004; 2004EP-00002722.
XX
PA (SELE-) SELEXIS SA.
XX
PI Mermod N, Girod PA, Bucher P, Nguyen D, Calabrese D, Saugy D;
PI Puttini S;
XX
DR WPI; 2005-333507/34.
XX
PT New purified and isolated DNA sequence having protein production
PT increasing activity comprises a bent DNA element and a binding site for a
PT DNA binding protein, useful for increasing protein production activity in
PT eukaryotic host cell.
XX
PS Disclosure; SEQ ID NO 53; 282pp; English.
XX
CC The invention relates to a purified and isolated DNA sequence having
CC protein production increasing activity comprising at least one bent DNA
CC element, and at least one binding site for a DNA binding protein. The
CC purified and isolated DNA sequence comprising a first and a second
CC isolated matrix attachment region (MAR) nucleotide sequence, which is a
CC MAR nucleotide sequence selected from a purified and isolated DNA
CC sequence above, a purified and isolated MAR DNA above, a purified and
CC isolated cLysMAR element and/or fragment, a synthetic MAR sequence, a
CC sequence complementary to it, its molecular chimera, or its combinations
CC and variants, is useful for increasing protein production activity in a
CC eukaryotic host cell. The present sequence represents a human chromosome
CC 1 contig DNA.
XX
SQ Sequence 828 BP; 414 A; 9 C; 18 G; 387 T; 0 U; 0 Other;

Query Match 5.4%; Score 69.4; DB 14; Length 828;
Best Local Similarity 44.7%; Pred. No. 0.001;
Matches 351; Conservative 0; Mismatches 431; Indels 3; Gaps 2;

```
Qy        256 TATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTT 315

Db        807 TATATTATATAATGTATATAATTATATATTATATACATTATATAATATGTAATTATATAT 748

Qy        316 TATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAG 375

Db        747 TATATACAATGTTATATATTATATGATATATAATTACATATTATATATACTATATTATAT 688

Qy        376 GACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGA 435

Db        687 ATAATATATAATTATATATTATATATATTATATAATATAATTATATATTATATATATTAT 628

Qy        436 GCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAAC 495

Db        627 ATAATATAATTATATATTATATATATTATATAATATAATTATATATTATATATATTATA- 569

Qy        496 GCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCA 555

Db        568 TAATATAATTATATATTATATATATTATATAATATAATTATATAATATAATTATATATTA 509

Qy        556 GTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGC 615

Db        508 TATATATTATATAATATAATTATATATTATAATTATATATTATATATAATATAATATAAT 449

Qy        616 TATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATA 675

Db        448 TATATATTATAT--ATACTATATTATATATAATTATATTATATTATATCTATTATATATA 391

Qy        676 GCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTT 735

Db        390 ATTATATATTATATATACTATATTATATAATAATTATATATATTATATCTATTATATATA 331

Qy        736 CAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACA 795

Db        330 ATAATATATACTATAGTATATAATATATAATTATATATTATATATATTATAGTATATAAT 271

Qy        796 GTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTA 855

Db        270 TATATATATATTATATATTATATATACTATATTATATATTATATATAATTATATATTATA 211

Qy        856 TGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATA 915

Db        210 TATACTATATTATATATATTATATATATTATATATATAATATATATATAATATATA 151

Qy        916 GGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTA 975

Db        150 CTATATTATATATATTATATATAATTATATATTATATATACTATAGTATATATTATATAT 91

Qy        976 TTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTA 1035

Db         90 ATTATATATACTATAGTATATATTATATATATTATATATACTATATAACTATATATTATA 31

Qy       1036 TCTTT 1040

Db         30 TAATT 26
```

RESULT 13
AAL60729/c
ID AAL60729 standard; DNA; 220 BP.
XX
AC AAL60729;
XX
DT 03-SEP-2003 (first entry)
XX
DE Yersinia pseudotuberculosis mutant O-antigen transporter (5D12) gene.
XX
KW Infection; tuberculosis; veterinary treatment; antimicrobial drug;
KW vaccine; O-antigen transporter; Wzx; mutant; ds.
XX
OS Yersinia pseudotuberculosis.
OS Synthetic.
XX
PN WO2003044047-A2.
XX

PD 30-MAY-2003.
XX
PF 18-NOV-2002; 2002WO-GB005212.
XX
PR 19-NOV-2001; 2001GB-00027657.
XX
PA (UNLO) LONDON SCHOOL HYGIENE & TROPICAL MEDICIN.
XX
PI Karlyshev AV, Wren BW;
XX
DR WPI; 2003-457588/43.
XX
PT New virulence peptides encoded by a gene of Yersinia pseudotuberculosis,
PT useful for treating or preventing a condition associated with infection
PT by Yersinia, e.g. tuberculosis, particularly in veterinary treatment.
XX
PS Claim 1; Page 127; 130pp; English.
XX
CC The present invention relates to virulence genes of Yersinia species and
CC proteins encoded by such genes. Sequences of the invention are useful in
CC the manufacture of a medicament for treating or preventing a condition
CC associated with infection by Yersinia e.g. tuberculosis particularly in
CC veterinary treatment. They are useful in screening assays for identifying
CC antimicrobial drugs. They are also used to prepare vaccines. The present
CC sequence is Yersinia pseudotuberculosis mutant O-antigen transporter, Wzx
CC gene
XX
SQ Sequence 220 BP; 89 A; 36 C; 37 G; 58 T; 0 U; 0 Other;

Query Match 5.3%; Score 69; DB 8; Length 220;
Best Local Similarity 58.5%; Pred. No. 0.0011;
Matches 120; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
Qy        644 ATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTT 703
              | | |  || ||  | |     |  |  ||   | ||| | ||| |  | || || |||
Db        209 AGCTTCGTAGAATTAGATCTGATATTATTAAGCCTATTGTCCGTCGTGCCTCTGGTTTTC 150

Qy        704 TTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAA 763
              |  ||||| || | ||| | | | |  |  |||| || ||||| || || || || |||
Db        149 TAATTTTTTCTATTTTAACTACAATTATTATTCAGACTGATTACATAGTAATGTCCCAAG 90

Qy        764 GGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCT 823
              |||  |||  ||| ||  |  |||||  ||    ||||| ||||||||| | || |
Db         89 TTTTAAAACCTAGTGAGATAATAACATATAGCGTTTTAATGAAGATTTTTGGTATTGTAT 30

Qy        824 TTTTTATTTATACTGCTATTTTGCA 848
              | ||| | |   ||||  |  | ||
Db         29 TCTTTGTCTTCTCTGCGTTGCTTCA 5
```

RESULT 14
ABL33641
ID ABL33641 standard; DNA; 5449 BP.
XX
AC ABL33641;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1614.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1614; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 5449 BP; 1146 A; 90 C; 1299 G; 2914 T; 0 U; 0 Other;

Query Match 5.3%; Score 69; DB 6; Length 5449;
Best Local Similarity 44.7%; Pred. No. 0.0014;
Matches 377; Conservative 0; Mismatches 445; Indels 21; Gaps 2;

```
Qy        296 TTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTT 355
              | |||||  ||  || ||||| |||||||| |    | ||||    |  || |  |  ||
Db       3572 TATTTTTAGTTTTTTGATTTTGTATTTTTTTTTTTTTTTATTTGTTTTTGATTTTATGTTT 3631

Qy        356 CTTTTCATGAGGTATTA--CAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGT 413
              ||||  ||   |  |||  || | |  |  |   || | ||||||   |  | | | |
Db       3632 TTTTTTATTTAGGTTTATTTAGAAGATTATTATTTTGGTTTTTTTTTTTATAGTTTTTAGG 3691

Qy        414 TTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGG 473
              || ||  | |  |    | ||  || ||  |||   |||   | | |  |
Db       3692 GAGAGATTTTAGAAGGAGGGATTTTTTTTTTTTTATTATATAGTTTTGGGGAATTTTTAA 3751

Qy        474 GTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATA 533
              || |   |   | ||       | |  |||  |   || |          | | |
Db       3752 TTGTATGTTTTTAGTTTTGGTTTTTTAAAGTTAGTTTCGAGTTGGAGAAATGTGGTTGGG 3811

Qy        534 TATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCC 593
              | ||  |||     | |     ||  | |||| ||||   ||    |  ||  | |
Db       3812 GAAATGGTATGATTAGGTTGAGTTTATCGTAAATTTATGGTTATATATTTTGGGTTTGTT 3871

Qy        594 AGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAA 653
              |   | || |    ||||  | |    | | | | ||  |   |  ||| |  | ||
Db       3872 TTTAATTTTGTTAGTTTTGTAGTGTGGATTTTTTTGGTAGGTTTATTTTTTTATTTTTTG 3931

Qy        654 AACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTAC 713
              || |  |  || | ||  | |     ||||   ||||| |||  | ||||||  | |
Db       3932 AAAATAATTATTTTATGTTTTTTTTTTTTTTTTTAGATTTTCGCGTTTTTTTTTATGAGT 3991

Qy        714 TTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCC 773
              ||| |  | ||| ||  |  || |   ||      ||  | || | |
Db       3992 TTTTTAGTTGATGGTTTTTTTTGAGGAAGGAAGAATTATTATATAT---------------- 4037

Qy        774 TGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTA 833
              |||||||| ||| || || |     |      ||||  ||   | ||||||||||
Db       4038 -----ATATTGTTAAATTTATAGAATTATTTTTTTATTTTATTTCGTTTTTTTTTATTTT 4092

Qy        834 TACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATG 893
              |   |  ||| | || ||  ||  | |  |          ||       ||   |
Db       4093 GTTAGAATAATAATAAGAAGTAAGGGTTAGGTTTTTTTGGGGGTTTAGGATTTTAGTTTTT 4152
```

```
Qy        894 GAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGG 953
              ||||  |  |  |  |      | ||||  |||  |   |||  | || ||
Db       4153 TTTGTTTTTTAAAGATTTTTTTAGGTGTTTTTTTTTTATTGTTTATTTATTTTTTTTTAT 4212

Qy        954 ATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATAT 1013
               |||  |   || |   || ||||    | | ||| |   |||   |    |||    |
Db       4213 TTGTGAATAGATTTTAGTGTTTTTTATTTTAAAAATAATGGTAGTGGTGATAAAATTATT 4272

Qy       1014 TAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGT 1073
              |  ||        |  | |||| | |||  ||||   |     | | ||| | |||
Db       4273 TTTTTGGTTTTATATTTTTTTGTATTTATTGTTTTATTTTTTTTTTTTTTTTATAGTAAA 4332

Qy       1074 TTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGAT 1133
               |  | ||  |  | |  |   |  |||     |  |   ||||||    | |  |||
Db       4333 ATTTTTTGGAAAGTGTTTATATTAGTTGTTTTGATTTTTTTATTTTTTATTTTTTTTTTA 4392

Qy       1134 ATT 1136
              |||
Db       4393 ATT 4395
```

RESULT 15
ACA64845
ID   ACA64845 standard; DNA; 115218 BP.
XX
AC   ACA64845;
XX
DT   27-JUN-2003 (first entry)
XX
DE   Human HNRNP GP43 DNA corresponding to AL034397.
XX
KW   Human; chronic inflammatory joint disease; infection; tumour;
KW   antiinflammatory; cytostatic; antiarthritic; antirheumatic;
KW   immunosuppressive; gene therapy; etiological pathogenicity; ds.
XX
OS   Homo sapiens.
XX
PN   DE10127572-A1.
XX
PD   05-DEC-2002.
XX
PF   30-MAY-2001; 2001DE-01027572.
XX
PR   30-MAY-2001; 2001DE-01027572.
XX
PA   (PATH-) PATHOARRAY GMBH.
XX
PI   Haeupl T, Ungethuem U, Blaess S;
XX
DR   WPI; 2003-240797/24.
XX
PT   Reagents for diagnosis, study and therapy of chronic inflammatory joint
PT   and other diseases, comprises any of many specified genes or derived
PT   proteins.
XX
PS   Claim 1; Page; 12pp; German.
XX
CC   This invention describes a novel reagent for diagnosis, molecular
CC   definition and therapy of chronic inflammatory joint diseases, and other
CC   inflammatory disorders, infective or tumour diseases in humans. The
CC   products of the invention have antiinflammatory, cytostatic,
CC   antiarthritic, antirheumatic and immunosuppressive activity and can be
CC   used for gene therapy. The reagent of the invention and any proteins and
CC   antibodies derived from it, are used (i) for analysing tissue and blood
CC   samples for medical diagnosis; (ii) for diagnosis and characterisation of
CC   chronic joint diseases, on the basis of molecular characterisation, and
CC   determining the etiological pathogenicity principle of as yet
CC   uncharacterised inflammatory diseases, also monitoring progression and/or
CC   treatment of disease, and optimisation of therapy and (iii) for
CC   developing treatments for inflammatory diseases, particularly of joints,
CC   infections and tumours. ACA64801-ACA64965 represent human polynucleotides
CC   used in the method of the invention

XX
SQ   Sequence 115218 BP; 35618 A; 23477 C; 22286 G; 33837 T; 0 U; 0 Other;

  Query Match   5.3%; Score 68.2; DB 8; Length 115218;
  Best Local Similarity   45.5%; Pred. No. 0.0024;
  Matches 522; Conservative   0; Mismatches 603; Indels   22; Gaps   7;

```
Qy         53 CATGGTTATCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTT 112
              | | ||  ||     |   ||| ||  | |      |||     |  || | || |  ||| |
Db      16487 CTTAGTAATAAGTATATATACTTATCAGTAATAAGTATAATATAATATAATGTTATATAT 16546

Qy        113 CTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTT-TGTTAACTGGTTTATTAGTATGG 171
                 || ||  |  | | ||   | | | || || |  ||  |  | ||| ||||
Db      16547 TATATATAAGTAATTATCAGTAATAATTATATAATTATTATATATTATATATAAGTAATT 16606

Qy        172 TGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGA 231
               ||    || ||  |   |||     || | |  || || ||| |    |  | |
Db      16607 ATCAGTAAAGTGGAAGTAATAATAAAGGAAGTAATATAATATAATAAGTATATAAATATA 16666

Qy        232 GCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCT 291
                | || | ||  ||| ||| |||||||| || |   |   ||   |  |       |
Db      16667 TAATATAATATAATATATTATGAATATATTATATTATATTATTATTATATAATATATATT 16726

Qy        292 ATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTT 351
              ||   || | ||||   |||| |||| || | |      ||      |||| ||
Db      16727 ATATATTGTATTATATAATTATTATATATATTATATTATTATATAATTATATTATATATT 16786

Qy        352 TCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTG 411
               |  ||  |||| || ||    | |  |  |  ||    |  | ||  | || | |
Db      16787 ATATTATTATATGATATA-TATTATATATTAATATTATTACATAATATATATTATATATT 16845

Qy        412 GTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACT-TTTTGCCGAATTGGT 470
               | ||   || |  | |||      |||   |||||  ||| | || |      ||   |
Db      16846 ATATTATATTATTATATAATATATTATATATTATATTATATTATTATATAATATATAT 16905

Qy        471 CGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCT 530
               |  |   || | ||| | |   ||||| |  |||  ||  |             |
Db      16906 TATATATTATATTATTATATAATATATATTATATATTATATTATATTATTATATAATATA 16965

Qy        531 ATATATATACTATAGGGGGATCTCAGTTGACATA------AAATTATCACTAATAGTCCT 584
              ||||||  ||||      | | |  | | |||       | |||||     ||  |  |
Db      16966 TATTATATATTATATTATATTATTATATAATATATATTATATATTATATTATATTATTAT 17025

Qy        585 GTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTA 644
               |||  |  | |   ||| || | ||  |  |||||  ||||||          |  ||
Db      17026 ATATTATATATTATATATTATATTATTATTATATAATATATATTATATATTATATTAT 17085

Qy        645 TCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTT 704
               |  | | |   || | ||  |||| | | |   || || ||    | | |
Db      17086 ATTATTATATAATATTTTATATTATATTATATTATTATATAATATATATTATATGATATA 17145

Qy        705 TCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAG 764
              |  | ||| | ||| ||  ||| |       |  |  | |||| ||  |   || | |
Db      17146 TTATATTATTATATAATATATTATATGATATATTATATTACATTATATATTATATTA- 17204

Qy        765 GCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTT 824
                  ||||||  |  |  ||| | ||       |  || | |  | |  | ||
Db      17205 ----------TATGATATATTATATGATATATTATATTATATGATATATTATATTATATA 17254

Qy        825 TTTTAT-TTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCA 883
              || |||  |||| |  || |  ||   | |||  | ||||       |  | | |
Db      17255 TTATATAATATAATATAATATATCATATAATATATCATATAATATAATATATCATATAAT 17314

Qy        884 AACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTAT 943
              |       ||   |||| |  |  | |  |||       |||| ||  |      |  ||
Db      17315 ATATTATATAATAAAATATATAATATAATATATAATATAATATATTATTATATATATAAT 17374

Qy        944 ATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAG 1003
              || || |   |  ||  |||| |||     ||| |||          || |    || ||
Db      17375 ATATTATATAACATATTATATATATAATAATATATAATATTATATAATATATTATATTAT 17434

Qy       1004 CCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCT 1063
                | |       || ||  |  || |||| |||  | |||  | | |    |||| || |
```

```
Db      17435 ATATAATATAATATAATATATATTATATTATATTATATATTATTAATATATTATAATTAT 17494

Qy       1064 GT-ATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAA 1122
              | ||     | |  | | |   | ||  | |  || |  |  ||||  |  ||| | | |
Db      17495 ATAATATATATATTATATAATTATAATATATTAATAATATATCGTATTATATATATTATA 17554

Qy       1123 ATACTTTGGATATTAGTACCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTT 1182
              ||  ||  ||||||        | |    ||   |  |  ||    |  ||  ||| ||
Db      17555 TTATATTATATATTATATTATTAAAATATTATCTATATTATAAATATATTATATTATATT 17614

Qy       1183 TCTAGTA 1189
              | || ||
Db      17615 TTTATTA 17621
```

Search completed: May 11, 2006, 09:43:00
Job time : 593.159 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 09:43:19 ; Search time 4028.56 Seconds
(without alignments)
15016.694 Million cell updates/sec

Title: US-09-423-093B-4_COPY_12762_14054
Perfect score: 1293
Sequence: 1 gtgaaagttcaattgttaaa.....ttaaggcaaataagggataa 1293

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	96.6	7.5	1272	1	AJ927522	AJ927522 AJ927522
c	2	90.8	7.0	1380	1	AJ928744	AJ928744 AJ928744
c	3	90	7.0	1060	10	CW937448	CW937448 TcB16.2_H
	4	87.4	6.8	1392	10	CG757503	CG757503 P052-4-C0
	5	86.4	6.7	2157	10	CL081966	CL081966 CH216-165
c	6	85.6	6.6	969	10	CW953164	CW953164 TcB39.1_G
c	7	84.4	6.5	1101	10	CNS00EVL	AL069706 Drosophil
c	8	84	6.5	1608	10	CL118721	CL118721 ISB1-72J8
c	9	84	6.5	1745	10	AG338221	AG338221 Mus muscu
	10	83.4	6.5	1981	10	CL082000	CL082000 CH216-165
c	11	83.2	6.4	1493	10	CL078589	CL078589 CH216-151
c	12	81.2	6.3	1632	10	CL082569	CL082569 CH216-167
c	13	80.8	6.2	1042	10	CL461494	CL461494 SAIL_1148
	14	80.2	6.2	1506	10	AG278469	AG278469 Mus muscu
c	15	79.6	6.2	1239	10	CW953278	CW953278 TcB39.2_C
	16	79.6	6.2	1436	8	DN807478	DN807478 76953322
	17	79.6	6.2	1542	10	AG386981	AG386981 Mus muscu
	18	79.6	6.2	1811	10	CG753732	CG753732 P048-4-G0
	19	79.4	6.1	1566	10	CG757757	CG757757 P053-1-D0
	20	78.6	6.1	1101	10	CNS00EVL	AL069706 Drosophil
c	21	78.4	6.1	1260	10	CL491610	CL491610 SAIL_559_
	22	78.4	6.1	1594	10	CL110653	CL110653 ISB1-53P2
	23	78.2	6.0	1378	10	AG350209	AG350209 Mus muscu
	24	78	6.0	1201	10	CNS0167M	AL106396 Drosophil
c	25	77.8	6.0	886	8	DN567115	DN567115 93902841
c	26	77.2	6.0	1605	8	DN712410	DN712410 CNB02-B08
c	27	77	6.0	996	10	CNS00FUH	AL071063 Drosophil
c	28	77	6.0	1238	1	AJ925855	AJ925855 AJ925855
c	29	76.6	5.9	1314	10	CL077082	CL077082 CH216-143
	30	76.2	5.9	1210	10	CG749728	CG749728 P044-1-C0
	31	76.2	5.9	1248	8	DN808090	DN808090 73970127
	32	76	5.9	1612	10	AG320745	AG320745 Mus muscu
	33	75.8	5.9	1251	10	AG332167	AG332167 Mus muscu
	34	75.8	5.9	1283	10	CG749744	CG749744 P044-1-C1
c	35	75.8	5.9	1364	8	DN693282	DN693282 CGX87-B03
	36	75.6	5.8	1015	10	CL135318	CL135318 ISB1-106F
	37	75.6	5.8	1074	9	BZ696936	BZ696936 SP__Ba009
c	38	75.6	5.8	1512	10	CL082685	CL082685 CH216-169
	39	75.6	5.8	1805	10	CL080711	CL080711 CH216-159
	40	75.6	5.8	1896	10	CG753083	CG753083 P048-1-C0
c	41	75.4	5.8	1416	10	CL499036	CL499036 SAIL_662_
	42	75.4	5.8	1473	10	AG278179	AG278179 Mus muscu
c	43	75.2	5.8	1536	10	CL078538	CL078538 CH216-151
	44	75	5.8	949	6	CA465606	CA465606 AGENCOURT
c	45	75	5.8	1162	10	CL077122	CL077122 CH216-143

ALIGNMENTS

RESULT 1
AJ927522/c
LOCUS AJ927522 1272 bp mRNA linear EST 10-JUN-2005
DEFINITION AJ927522 Theileria annulata merozoite Theileria annulata cDNA clone tam007c01_q1k, mRNA sequence.
ACCESSION AJ927522
VERSION AJ927522.1 GI:67497906
KEYWORDS EST.
SOURCE Theileria annulata
ORGANISM Theileria annulata
Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae; Theileria.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Pain,A., Renauld,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W., Kerhornou,A., Aslett,M., Bishop,R., Bouchier,C., Cochet,M., Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Fosker,N., Gardner,M., Goble,A., Griffiths-Jones,S., Harris,D.E., Katzer,F., Larke,N., Lord,A., Maser,P., McKellar,S., Mooney,P., Morton,F., Nene,V., O'Neil,S., Price,C., Quail,M.A., Rabbinowitsch,E., Rawlings,N.D., Rutter,S., Saunders,D., Seeger,K., Shah,T., Squares,R., Squares,S., Tivey,A., Walker,A.R., Woodward,J., Dobbelaere,D.A.E., Langsley,G., Rajandream,M.-A., McKeever,D., Shiels,B., Tait,A., Barrell,B. and Hall,N.
TITLE The genome of the host-cell transforming parasite Theileria annulata and a comparison with T. parva
JOURNAL Unpublished (2005)
COMMENT Contact: Pain A
The Pathogen Sequencing Unit
The Wellcome Trust Sanger Institute
Genome Campus, CB10 1SA, UNITED KINGDOM
Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
FEATURES Location/Qualifiers
source 1. .1272
/organism="Theileria annulata"
/mol_type="mRNA"
/isolate="Ankara (clone D7)"
/db_xref="taxon:5874"
/clone="tam007c01_q1k"
/dev_stage="merozoite"
/lab_host="Bos taurus (cow)"
/clone_lib="Theileria annulata merozoite"
/note="country: Turkey:Ankara"
ORIGIN

```
Query Match             7.5%;  Score 96.6;  DB 1;  Length 1272;
Best Local Similarity   45.3%;  Pred. No. 2.9e-08;
Matches  467;  Conservative    0;  Mismatches  552;  Indels   11;  Gaps    3;

Qy        246 TTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTAT 305
              || | ||  || |||| | | |  |||| || | |  || ||     |   ||||||
Db       1259 TTTTTATTTATTTATTTATTTAATATTATATGTTATTTTAATTTAATTATGTTTTTTATA 1200

Qy        306 TGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCATGA 365
              |   ||| ||||| ||| || |    |  |||      ||| | ||| |   | |
Db       1199 TTTATTAATTTATTTTTATTATTAATTTTTTTATATGTAATTTATTTTTATATATTTATT 1140

Qy        366 GGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTAT 425
                |   |        |||  |   |  | || |||| | || | |   |||   || |||
Db       1139 TTTTANATTTATTTAAATTAATTTTTTATTATTTATTTTATTTATTTATTTTTATAT 1080

Qy        426 TGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAA 485
              |   ||      | ||  ||  || |||  || |   |  |              |  |
Db       1079 TATTATATTTATTTTTTATTTAAATATATTATATTTATATTTATTAAATTTTATTTTT 1020

Qy        486 TCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTATAG 545
              | ||||      |||||  |  |  |  | |       |    | |||||  | |
Db       1019 TATATTTTTATTATTATTATTATTTTTTTATTTTTATATTATTTATTTATATTATTTTA 960

Qy        546 GGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGAT 605
                |  |   ||     |  ||||   | | ||  |  | | |  |    |   ||| ||
Db        959 TTATTTTTTTTTTTTTTTTTAATTTATATTTATTATATTATTTTATTTTTTTTTTATTAT 900

Qy        606 TTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATC 665
              ||  ||  |     | ||   | || ||    |      |||||  | | | |      | ||
Db        899 TTATTTTTTTATTTTTTATTTTATATTTTTTTTATTTTATATTTTAATATTTTTTATATT 840

Qy        666 TCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGAT 725
              |   | |||   |||||   |  || ||     | || | |||||  ||||| | |   |
Db        839 TATNTTTATTTTTATTTTTTTTTTTATTTTTTATTTATTATATTTTTTTTTTATTTTTTT 780

Qy        726 AGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGT 785
              | |  |  ||  |      || |||    ||||  |    |   |  |  |  |||| |
Db        779 ATTTTTTATTATATATTTTTTTAT----ATTTATTTTATTTTTTATTTTTTTATTATTTT 724

Qy        786 TCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTT 845
              |   | |  | | |   |   |||||||  || |   || |||| || |  |    | ||
Db        723 TTTTTTTTTATTTATTTTTTAATTTTTTATTAATATTTATTTAATTTTTATTTATTATT 664

Qy        846 GCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAA 905
                  ||| |   | | | |  | ||||    | |    || | ||  |
Db        663 TATTATATTTTTAAATTTTTTTTTTAAATTTTATTTTA------TTTGAATAATTTTATT 610

Qy        906 CAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATT 965
                |  | |     |||||  ||   | |      | || ||  || |    |||| |
Db        609 TTTTTAANTTTTTATTTTAATTTTTTTTTTATAAAATATTTTTATTTAATATTTTTTATT 550

Qy        966 TATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGT 1025
              || || ||| |||      |   | | ||      |||        | ||  || | |  |
Db        549 TAATTTTTTTTTTTTTTTATATATTTTATTTTTATATATATTTTTTTTTTTTTTTTATAT 490

Qy       1026 TTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACAC 1085
              ||  ||||||| |||| | ||  ||   | || ||| | ||||   |||  |  | | |
Db        489 TTAAATTTTATTTTTTTTTTTTTTTTTTTAAATTTTTTTATATTTATTTTATTTAAAAAA 430

Qy       1086 TTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTACCACT 1145
              || |  ||| ||||        |||  |||   |||| |    |   | | | ||    |
Db        429 TTTTTAAAT-TTATATTTTTTTATTTATTTGATAAATTTTTAGTTTTTTTAAATATTTTT 371

Qy       1146 ACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATCAGTGG 1205
                |    |  ||| |  ||||  |    |  ||||||| || |    ||   ||   |
Db        370 TATAAATTTTTTTGTTTTAATAAAATTTTAATATTTTTTTAATGAATTTTTTATTTTTTT 311

Qy       1206 AGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCACTAAC 1265
              |    ||      || ||||| | ||||  || || | |||| |      ||     ||
Db        310 ATATATGAAAAAATTAATTATTTTTTTTATTTAAATTATTTTATAAAATTTTTTTTTATA 251

Qy       1266 TTACTTAATT 1275
```

```
                ||  | ||  |
Db        250 GTATTAAAGT 241
```

RESULT 2
AJ928744/c
LOCUS AJ928744 1380 bp mRNA linear EST 10-JUN-2005
DEFINITION AJ928744 Theileria annulata merozoite Theileria annulata cDNA clone tam022c08_q1k, mRNA sequence.
ACCESSION AJ928744
VERSION AJ928744.1 GI:67499132
KEYWORDS EST.
SOURCE Theileria annulata
ORGANISM Theileria annulata
Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae; Theileria.
REFERENCE 1 (bases 1 to 1380)
AUTHORS Pain,A., Renauld,H., Berriman,M., Murphy,L., Yeats,C.A., Weir,W., Kerhornou,A., Aslett,M., Bishop,R., Bouchier,C., Cochet,M., Coulson,R.M.R., Cronin,A., de Villiers,E., Fraser,A., Fosker,N., Gardner,M., Goble,A., Griffiths-Jones,S., Harris,D.E., Katzer,F., Larke,N., Lord,A., Maser,P., McKellar,S., Mooney,P., Morton,F., Nene,V., O'Neil,S., Price,C., Quail,M.A., Rabbinowitsch,E., Rawlings,N.D., Rutter,S., Saunders,D., Seeger,K., Shah,T., Squares,R., Squares,S., Tivey,A., Walker,A.R., Woodward,J., Dobbelaere,D.A.E., Langsley,G., Rajandream,M.-A., McKeever,D., Shiels,B., Tait,A., Barrell,B. and Hall,N.
TITLE The genome of the host-cell transforming parasite Theileria annulata and a comparison with T. parva
JOURNAL Unpublished (2005)
COMMENT Contact: Pain A
The Pathogen Sequencing Unit
The Wellcome Trust Sanger Institute
Genome Campus, CB10 1SA, UNITED KINGDOM
Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
FEATURES Location/Qualifiers
source 1. .1380
/organism="Theileria annulata"
/mol_type="mRNA"
/isolate="Ankara (clone D7)"
/db_xref="taxon:5874"
/clone="tam022c08_q1k"
/dev_stage="merozoite"
/lab_host="Bos taurus (cow)"
/clone_lib="Theileria annulata merozoite"
/note="country: Turkey:Ankara"
ORIGIN

```
  Query Match             7.0%;  Score 90.8;  DB 1;  Length 1380;
  Best Local Similarity   41.0%;  Pred. No. 3.8e-07;
  Matches  350;  Conservative    0;  Mismatches  497;  Indels    7;  Gaps    2;

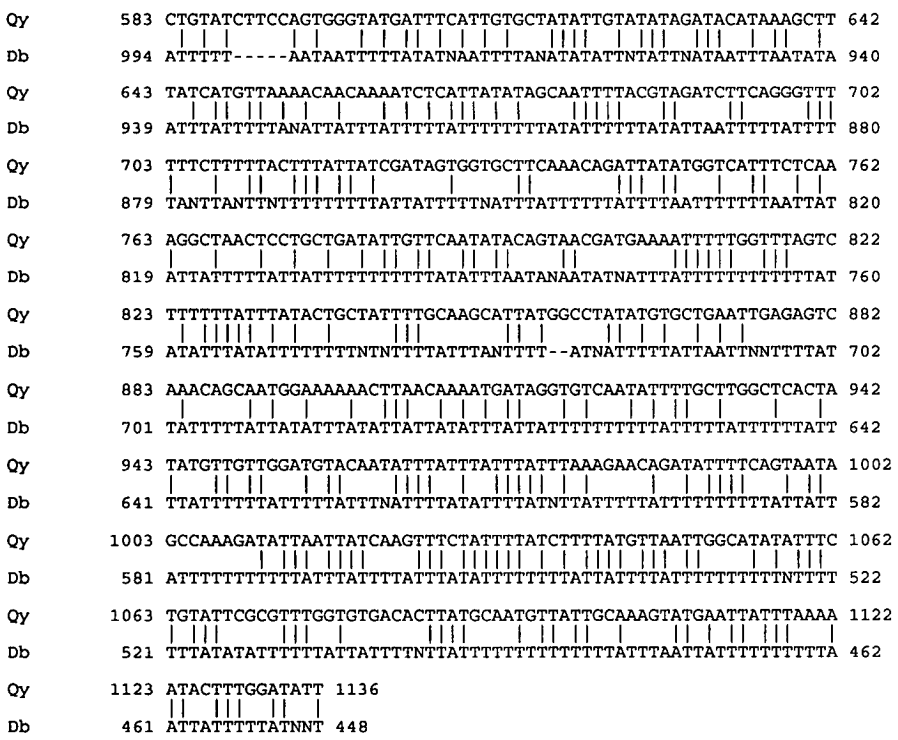
Qy        283 TTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCT 342
              ||| |   |  | ||||||||  ||  ||  |||| | |||| | |    |  |||    |
Db       1294 TTTTTNTTTTNTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTNTNTTTTTTTTTTTTTTT 1235

Qy        343 AAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACC 402
                | |  || ||  |||  |       |                              |    |
Db       1234 TTTTTTTNTTTTTNNTTTNTTNNNNNNNTNNNTNNNTNNNNNNTNNNNNTNNNNNTNNN 1175

Qy        403 TCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCC 462
                                |                              |||| | ||
Db       1174 NNNNNNNNNTNNNNNNNTNNNNNNNNNTNNNNNNNNNNNNNNNCCCCAATATTATTATTT 1115

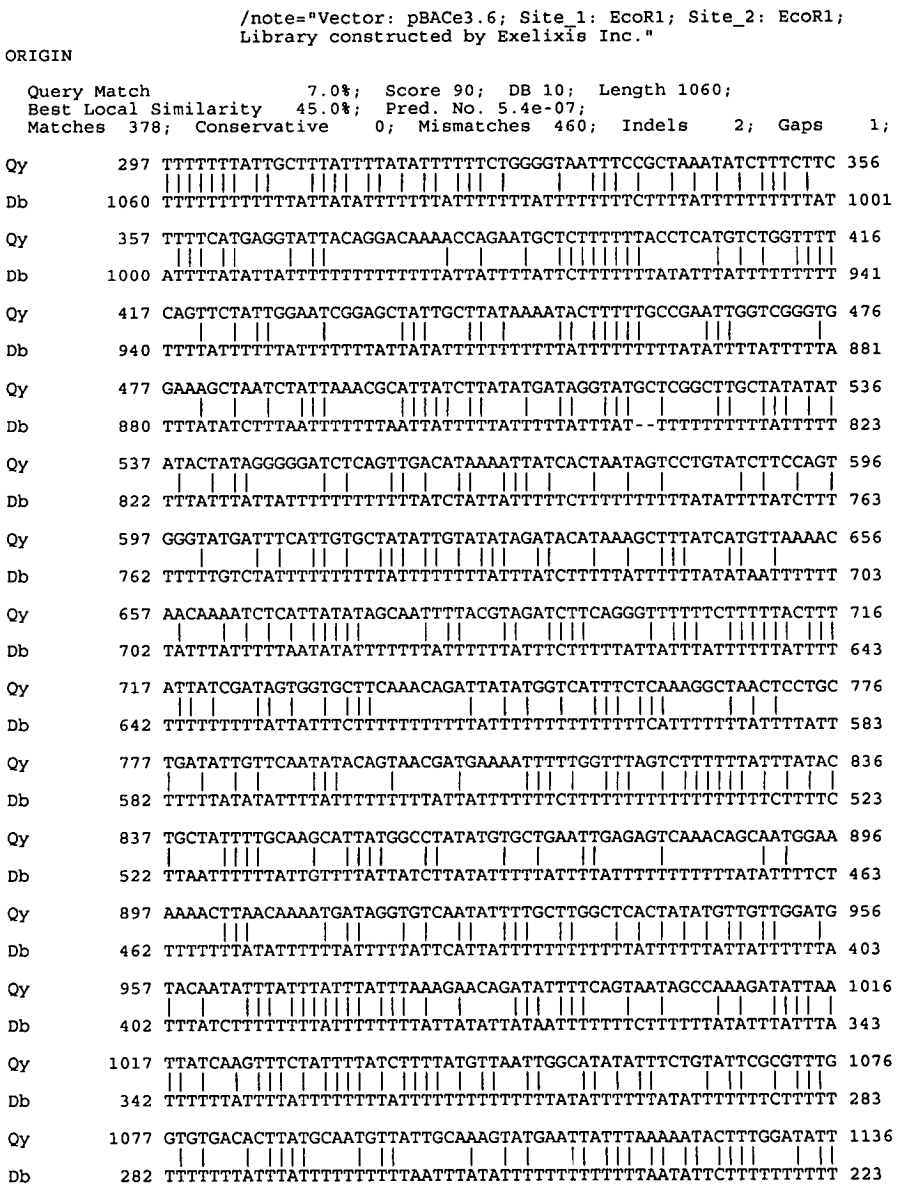
Qy        463 GAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTC 522
              ||||  |   |   |  || | ||||  |  || || |  | |  ||       |
Db       1114 TAATTTTTTTTTTATTATAATATTTTATTTTAATTATAATTTATTTTTTTANTATATTTN 1055

Qy        523 GGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTC 582
                ||  | ||| | |  |         ||      |    |||| ||| | | | |||
Db       1054 TANTTAATTTATTTTTTATTATTTTTTATTATTATANTTATAATATTNTTAATTATATAA 995
```



```
RESULT 3
CW937448/c
LOCUS       CW937448               1060 bp    DNA     linear   GSS 20-DEC-2004
DEFINITION  TcB16.2_H03_SP6 Tribolium BAC library Tribolium castaneum genomic,
            genomic survey sequence.
ACCESSION   CW937448
VERSION     CW937448.1  GI:56720061
KEYWORDS    GSS.
SOURCE      Tribolium castaneum (red flour beetle)
  ORGANISM  Tribolium castaneum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
            Tenebrionidae; Tribolium.
REFERENCE   1  (bases 1 to 1060)
  AUTHORS   Savard,J. and Tautz,D.
  TITLE     Tribolium castaneum BAC-ends sequencing project
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Savard, J.
            Abteilung fur Evolutionsgenetik, AG Tautz
            Institut fur Genetik, Universitat zu Koln
            Weyertal 121, 50931 Koln, Germany
            Tel: 49 221 470 6911
            Fax: 49 221 470 5975
            Email: savard@uni-koeln.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1060
                     /organism="Tribolium castaneum"
                     /mol_type="genomic DNA"
                     /strain="GA-2"
                     /db_xref="taxon:7070"
                     /clone_lib="Tribolium BAC library"
                     /note="Vector: pBACe3.6; Site_1: EcoR1; Site_2: EcoR1;
                     Library constructed by Exelixis Inc."
ORIGIN

  Query Match             7.0%;  Score 90;  DB 10;  Length 1060;
  Best Local Similarity   45.0%;  Pred. No. 5.4e-07;
  Matches  378;  Conservative    0;  Mismatches  460;  Indels    2;  Gaps    1;
```



```
RESULT 4
CG757503
LOCUS       CG757503               1392 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P052-4-C08.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG757503
VERSION     CG757503.1  GI:37986131
KEYWORDS    GSS.
```

```
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1392)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1392
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN

  Query Match             6.8%;  Score 87.4;  DB 10;  Length 1392;
  Best Local Similarity   44.2%;  Pred. No. 1.7e-06;
  Matches  447;  Conservative    0;  Mismatches  559;  Indels    6;  Gaps    2;

Qy        271 TTACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGG 330
Db        147 TTTAATTTTTTTTATAATAATATATTATTTTAAATTTTTTTTAAATTTTATTAAAAAA 206

Qy        331 GTAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATG 390
Db        207 TTAATATATATTTTTTTTTTTTTTTTTTATTTTAAATTTATTTAATATTTTTTATATATTT 266

Qy        391 CTCT---TTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTAT 447
Db        267 ATATAAATTTTTAATATATATAATTTATAAATTTATATATTTATAATAATAAATATAAAT 326

Qy        448 AAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTAT 507
Db        327 TAATTATATATTTTTTTTTTATTTTATATTATAATTTTTATTTTATTTATTTTTTATTTT 386

Qy        508 ATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAA 567
Db        387 AAATATAATATTTTATTATTATTTTTAATTTATTTTATTTATTTTTTTTTTTTATTATTA 446

Qy        568 TTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATAT 627
Db        447 TTTTTTTTAATTATATTTTTTTTTTTAATTATTTTTTTTTTTTTTTTTTATATATTATAT 506

Qy        628 AGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGT 687
Db        507 AAATATATATAATTAATAATTAATAAAATAAATAAAATTAATATAAATTTATTTAAAAAT 566

Qy        688 AGAT---CTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGAT 744
Db        567 TTTTATATTTTTTTTTTTTTATTTTTTTATTATANATTTTTTTTTATTTTTATAAATATTT 626

Qy        745 TATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATG 804
Db        627 TATTTTTTATTTTTATTTTTTTTATATTTTAATTTATTTATTTATATTTTATTTAATATT 686

Qy        805 AAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATA 864
Db        687 TTTATTTTATTTTTTTTTTTTTATATTTTTATTTTTTTTTTTATTTATTTTTTTTTTTA 746

Qy        865 TGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAAT 924
Db        747 TTTATTTAATAATTAATTATATTTATTTATATATATTTTTTTATTTTTTTTTTTATTTTTA 806

Qy        925 ATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAA 984
Db        807 TTATTTATTTTATTTATTTAATTAATATATATTTTTTAATTTTTTTTTTTTTTTTTTTAT 866

Qy        985 CAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATG 1044
Db        867 AATATAATTTATTATTTTATATTTAATTTTTATAATTTTATTTATTTTTNTTTTTATATT 926

Qy       1045 TTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAA 1104
Db        927 TTATATATATATTAATTTTTTTTTTTTTTTTTTTTTTTTTATATATTTTTTTATATTTTTA 986

Qy       1105 AGTATGAATTATTTAAAAATACTTTGGATATTAGTACCACTACAAGCAATAATTGGTGGA 1164
Db        987 ATTTATTTTTATTTTATTATTTATTTATTTTTTATTTTTTTTATATATTAATAATTTTAT 1046

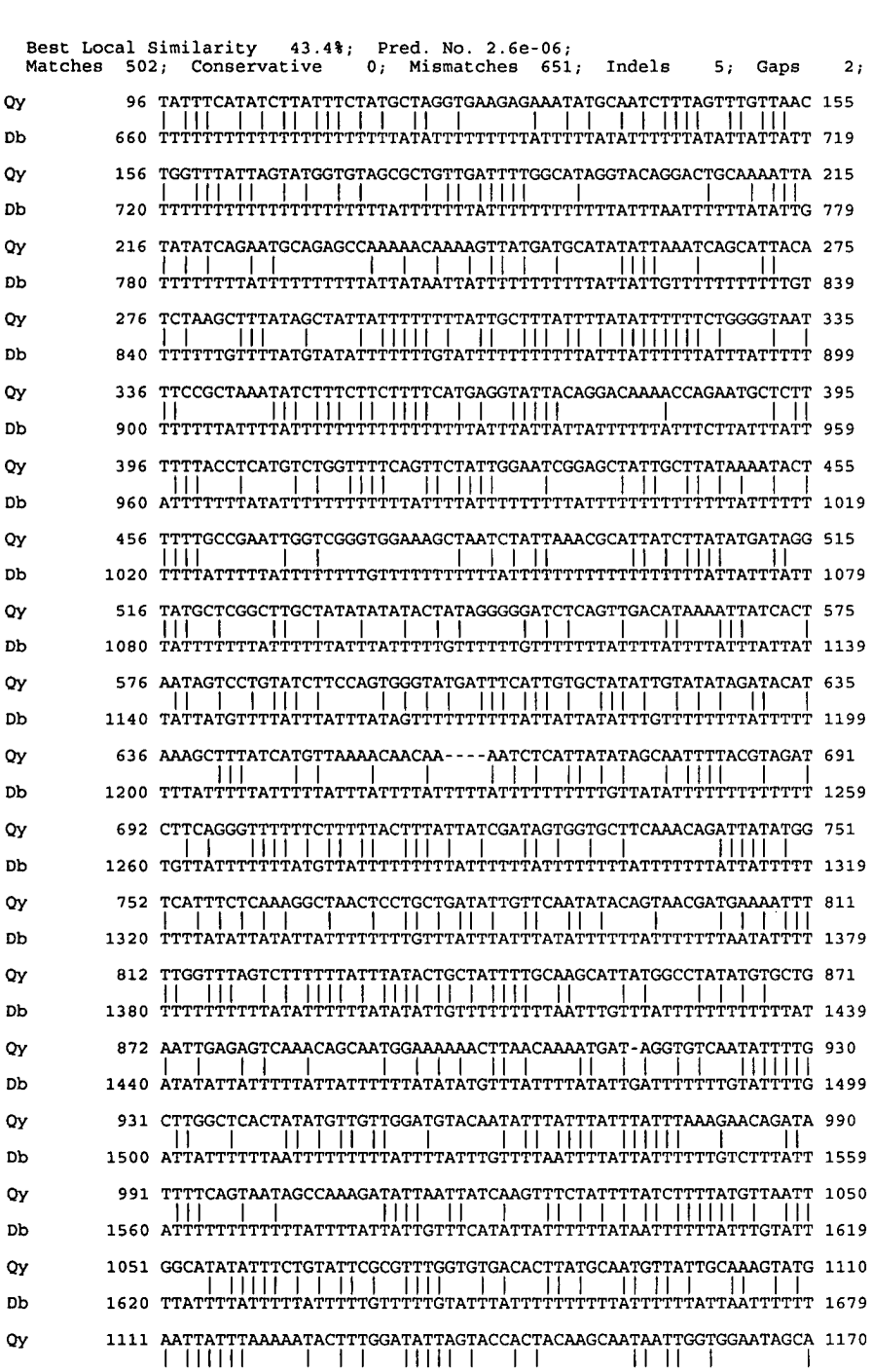
Qy       1165 ATAGCACAATGGTATTTTTCTAGTACGCTTGGAATCAGTGGAGTGCTGCTTGGCTTGATT 1224
Db        1047 TTAATATTTTTTTTATTTTATTATTAATTTTTATTTTTTTTATTATTATTTTTTTTATTT 1106

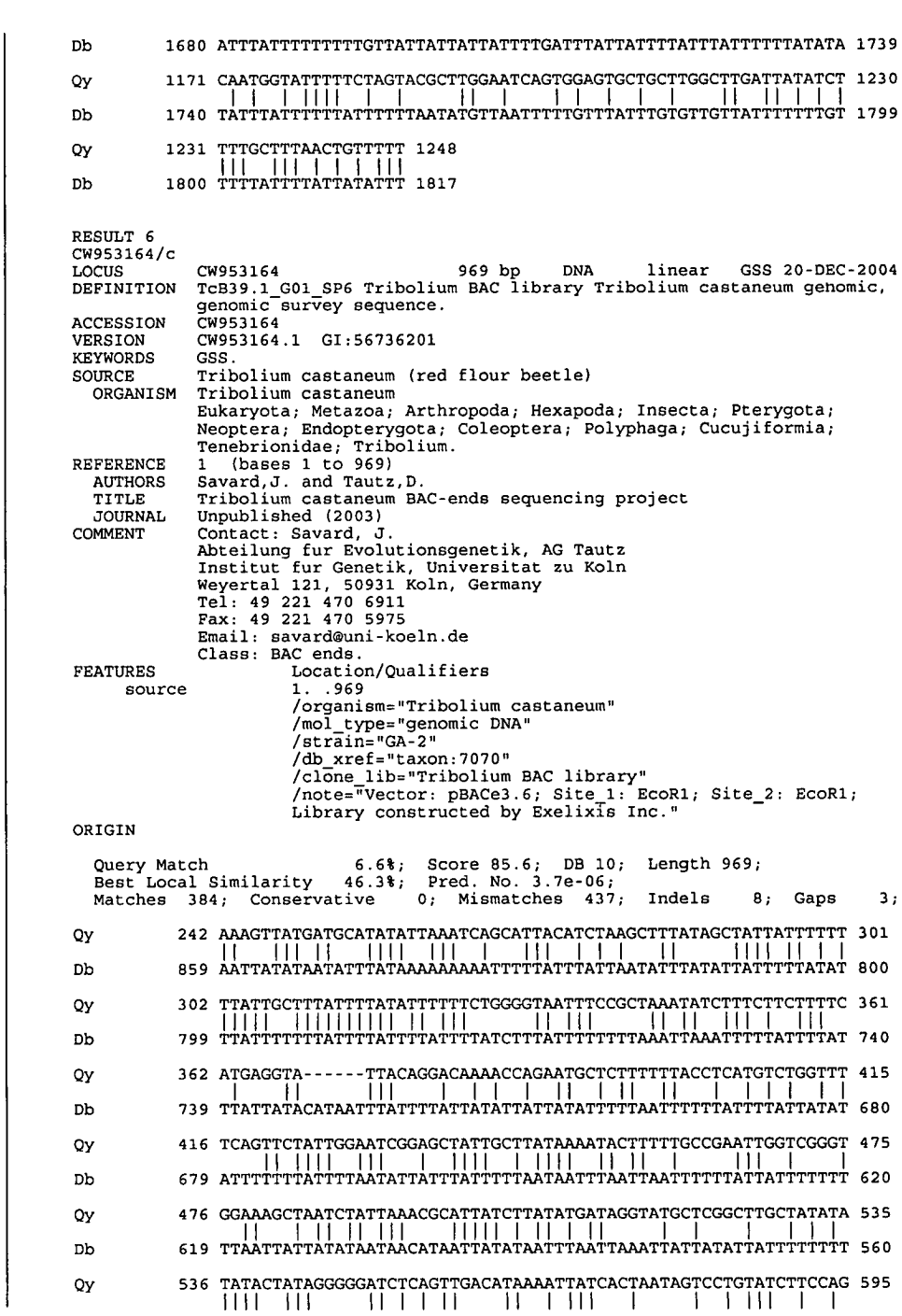
Qy       1225 ATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCACTAACTTACTTAATTA 1276
Db        1107 TTATTTTTATTTTTTATTTTTTTATATTTNATTTTTTATATTTATTTTAATA 1158


RESULT 5
CL081966
LOCUS       CL081966                2157 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-165D13_Sp5.1 CH216 Xenopus tropicalis genomic clone
            CH216-165D13, genomic survey sequence.
ACCESSION   CL081966
VERSION     CL081966.1  GI:40537879
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 2157)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp5 atctgccgtttcgatcct
            Class: BAC ends
            High quality sequence start: 341
            High quality sequence stop: 412.
FEATURES             Location/Qualifiers
     source          1. .2157
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-165D13"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             6.7%;  Score 86.4;  DB 10;  Length 2157;
```

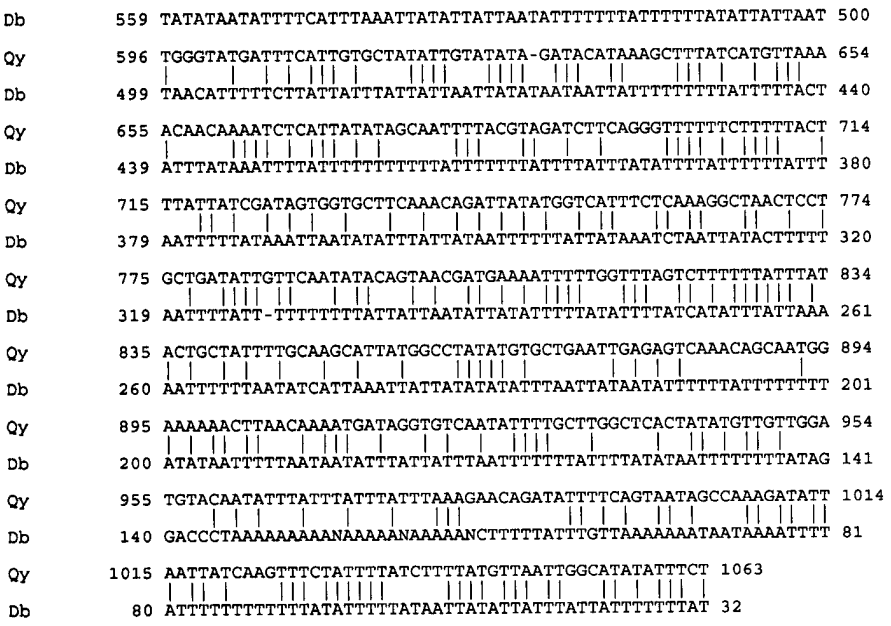
Best Local Similarity 43.4%; Pred. No. 2.6e-06;
Matches 502; Conservative 0; Mismatches 651; Indels 5; Gaps 2;





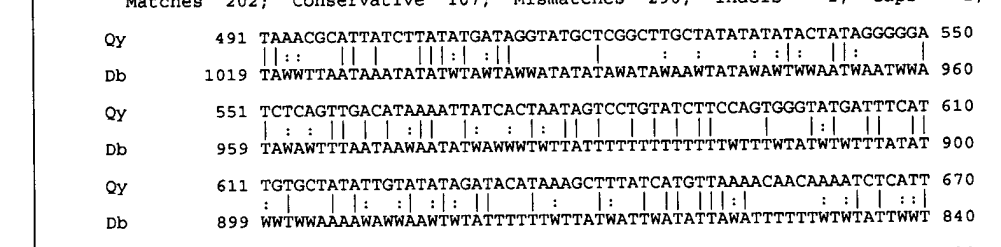
RESULT 6
CW953164/c
LOCUS CW953164 969 bp DNA linear GSS 20-DEC-2004
DEFINITION TcB39.1_G01_SP6 Tribolium BAC library Tribolium castaneum genomic, genomic survey sequence.
ACCESSION CW953164
VERSION CW953164.1 GI:56736201
KEYWORDS GSS.
SOURCE Tribolium castaneum (red flour beetle)
ORGANISM Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tribolium.
REFERENCE 1 (bases 1 to 969)
AUTHORS Savard,J. and Tautz,D.
TITLE Tribolium castaneum BAC-ends sequencing project
JOURNAL Unpublished (2003)
COMMENT Contact: Savard, J.
Abteilung fur Evolutionsgenetik, AG Tautz
Institut fur Genetik, Universitat zu Koln
Weyertal 121, 50931 Koln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: savard@uni-koeln.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .969
/organism="Tribolium castaneum"
/mol_type="genomic DNA"
/strain="GA-2"
/db_xref="taxon:7070"
/clone_lib="Tribolium BAC library"
/note="Vector: pBACe3.6; Site_1: EcoR1; Site_2: EcoR1; Library constructed by Exelixis Inc."
ORIGIN

Query Match 6.6%; Score 85.6; DB 10; Length 969;
Best Local Similarity 46.3%; Pred. No. 3.7e-06;
Matches 384; Conservative 0; Mismatches 437; Indels 8; Gaps 3;



RESULT 7
CNS00EVL/c
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"
ORIGIN

Query Match 6.5%; Score 84.4; DB 10; Length 1101;
Best Local Similarity 33.7%; Pred. No. 6.3e-06;
Matches 202; Conservative 107; Mismatches 290; Indels 1; Gaps 1;



RESULT 8
CL118721/c
LOCUS CL118721 1608 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-72J8_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-72J8, genomic survey sequence.
ACCESSION CL118721
VERSION CL118721.1 GI:40612356
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1608)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine

```
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 252
            High quality sequence stop: 345.
FEATURES             Location/Qualifiers
     source          1. .1608
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-72J8"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             6.5%;  Score 84;  DB 10;  Length 1608;
  Best Local Similarity   43.0%;  Pred. No. 7.5e-06;
  Matches  538;  Conservative    0;  Mismatches  700;  Indels   14;  Gaps    2;

Qy         33 TTTAATTGTTGCAGGTTCATCATGGTTATCCAAAATAATAATTGCCGGGGTGCAGTTAGC 92
Db       1485 TTTTTTTTTTTTTATTTTAATTTTATTTATATTTTATTTCTTTCTTTATTTTTATT 1426

Qy         93 AAGTATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTT 152
Db       1425 TATTTTTTTATTTTTTATAATTATTTTTTTTTATATTTATTTTATATTATTTATTTAAAT 1366

Qy        153 AACTGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAA 212
Db       1365 TTTTTTTTTATTTATTATTTAT-----TTTTTTTTTTTTCTAATATATTATTTTTTTTAT 1311

Qy        213 TTATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATT 272
Db       1310 ATTTAATTTTTATTTAATATTAATTTTAATTTTTATTTATTTTTTTTTTTTATATTTTTT 1251

Qy        273 ACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGT 332
Db       1250 ATTATTTTTTTATTTTATTTTTATTTTATTTTTTTTTTTTTTTTTTTTTTTTATTTTTTT 1191

Qy        333 AATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCT 392
Db       1190 AATTATTTATATTTTTTATTTATATTTTAATATTTTTTATTTTTTTAATATTTTTTATTT 1131

Qy        393 CTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAAT 452
Db       1130 TTATATAATATTTTATTTTATTTTAATTTTATTTTTTATTTTTTTTTTTTTATATTTATT 1071

Qy        453 ACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGAT 512
Db       1070 TTTATTTTTTTTCTTTTTTTTTTTTTATTTTAATTTTTAATTTTATTTTTATTTTTTTT 1011

Qy        513 AGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATC 572
Db       1010 TTTTTTTTTTTTTTTTTTTTTTATATTAATTTTTTATATTTTTTTTTTTTATTATATTTATT 951

Qy        573 ACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATA 632
Db       950 TTTTATTATATTTTTTTAATTTTTTTTTTTTATTATTATTTTTTTATATTATATTTTTTTT 891

Qy        633 CATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATC 692
Db       890 AATTTATTTAATTTTTTTTTTATTTTTTATTTTTTATATATTTTTTTTAATTTTTTTATT 831

Qy        693 TTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGT 752
Db       830 TTTATTTTTTTTTATTTTTTTTTTTTTTATTAATATTTTTATTTTATATTTATTTTTAATT 771

Qy        753 CATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTT 812
Db       770 TTTTTTTTTTATATATTATTTTTCTTATATATTTATTTTTATTATTTATATTTTTTATTT 711

Qy        813 TGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGA 872
Db       710 ATATTTTTTATTTATTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTAATATTTTT 651

Qy        873 ATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCT 932
Db       650 ATTTTTTTTTTTATTAATCTTATTATAAATTTTATTATTTCTTTTTTATATTATTTTTTT 591

Qy        933 TGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATT 992
Db       590 TATTTTATTTTTTAAAAAAATAATAATTGGGGTTTTTTTTTTTTTTTTTTTTCTCCCCTC 531

Qy        993 TTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGG 1052
Db       530 CAAAAAAAAAAAAAAAGTGGTTTTTTTTGCTTTTTTTTTTTTTTTTTTTTAAAAAAAAAAA 471

Qy       1053 CATATATTTCTGTATTCGCGT---------TTGGTGTGACACTTATGCAATGTTATTGCA 1103
Db       470 AAAATTTTTTTTTTTTTTTTTATATTTTTTTTATTTTATTTTTTTTTTATTTTTTTTTA 411

Qy       1104 AAGTATGAATTATTTAAAAATACTTTGGATATTAGTACCACTACAAGCAATAATTGGTGG 1163
Db       410 TATTTTAATTTATTTATATTTTTTTTTTTTATTTTTAATATTTTATTATTTATTTTTTAA 351

Qy       1164 AATAGCACAATGGTATTTTTCTAGTACGCTTGGAATCAGTGGAGTGCTGCTTGGCTTGAT 1223
Db       350 AATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 291

Qy       1224 TATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCACTAACTTACTTAATT 1275
Db       290 CTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTTTTTTTTT 239


RESULT 9
AG338221/c
LOCUS       AG338221               1745 bp    DNA     linear   GSS 18-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-130H15.T7, genomic survey
            sequence.
ACCESSION   AG338221
VERSION     AG338221.1  GI:47911531
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1745)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
```

```
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1745
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-130H15.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             6.5%;  Score 84;  DB 10;  Length 1745;
  Best Local Similarity   43.3%;  Pred. No. 7.5e-06;
  Matches  504;  Conservative    0;  Mismatches  656;  Indels    3;  Gaps    3;

Qy         12 ATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTATCCAAAATAAT 71
Db       1363 ATTATTACATATTCTAAGTATATTATCTATTATTTATATAT-ATAATTCTAATANATCAT 1305

Qy         72 AATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTAGGTGAAGAGAA 131
Db       1304 ATTTATATTTATATTTATATATTAATNTACATATATTCTATATAATTAATTTAATGAATA 1245

Qy        132 ATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGG 191
Db       1244 ATATTTATTATATTATCAATTATATTATTTTATATTTTATTATATTATTCTTAATTTTAT 1185

Qy        192 CATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGA 251
Db       1184 ACTATTTATATTTATTTTTAATTATATATATTATTATATTTATATTTTATTATATTTAAA 1125

Qy        252 TGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTT 311
Db       1124 TTATTATTTTATTACTTATTATTATATACATATATCTTATTATATTTTCATTTATATTAT 1065

Qy        312 ATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATT 371
Db       1064 ATATATTTTATTTTATTTTATTATTTATATTTATTTTTTTAATTTTTTTTTTTTATATTTT 1005

Qy        372 ACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAAT 431
Db       1004 TATATTATTTTTTATTTAAACTTTTTTTTATTATATATATATTATTATTTAAAAATAATT 945

Qy        432 CGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATT 491
Db        944 TTTTTTTTATATTAAAATATTATTTTCTTTTTATTATAATAATATTTTTTTTTATTTTTA 885

Qy        492 AAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGAT 551
Db        884 ATATACTATATTTAATTTATTTATTATTTTATTTAATTTTTTTATTTTTTTTATANATTA 825

Qy        552 CTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATT 611
Db        824 TTTATATTTAATTTTATATTTTATAATTTTTATTTCTTTATTATAATAAAATTTTTTTTT 765

Qy        612 GTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTA 671
Db        764 ATAATTTTTATTATATATATTTTTTTATATTTAT-TTTTAAATAAATTTTATTATTTTAT 706

Qy        672 TATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGT 731
Db        705 TTTATTATATTTATTTTTTTATTTTATTAATTTTATTTATTTATTATTTTAATATTTTTT 646

Qy        732 GCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATA 791
Db        645 ATAAATTATATATTATTTTTAAATTTATATTTTATTATATTTTAAATTATTATTTTATA 586

Qy        792 TACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTAT-TTATACTGCTATTTTGCAAG 850
```

```
Db        585 ATTTTATAAATTTTATTTTTATTATATATTTTTATATATATTTTAATTTTTTTTTTTTTT 526

Qy        851 CATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAA 910
Db        525 TTTTANAAAAANATTTTTATTATTAGGGGGGGGGGGGGAATAAAAATATTTTTTATTTAT 466

Qy        911 TGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTT 970
Db        465 TTATATTTTTATAAATTTTTTTANTTTTTNTTTTNNNTTAAAAAAAAAAAATATTTTTTT 406

Qy        971 ATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTA 1030
Db        405 TTTTTTTATTATTTTTTAATATTTTATTTAATTTAATTTATAATTTTTATTATATTTTTA 346

Qy       1031 TTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATG 1090
Db        345 TTTTTAAAAAAAATAAAATTTTAATATTATATAATTTTTTTTATTTATATTATATATTTT 286

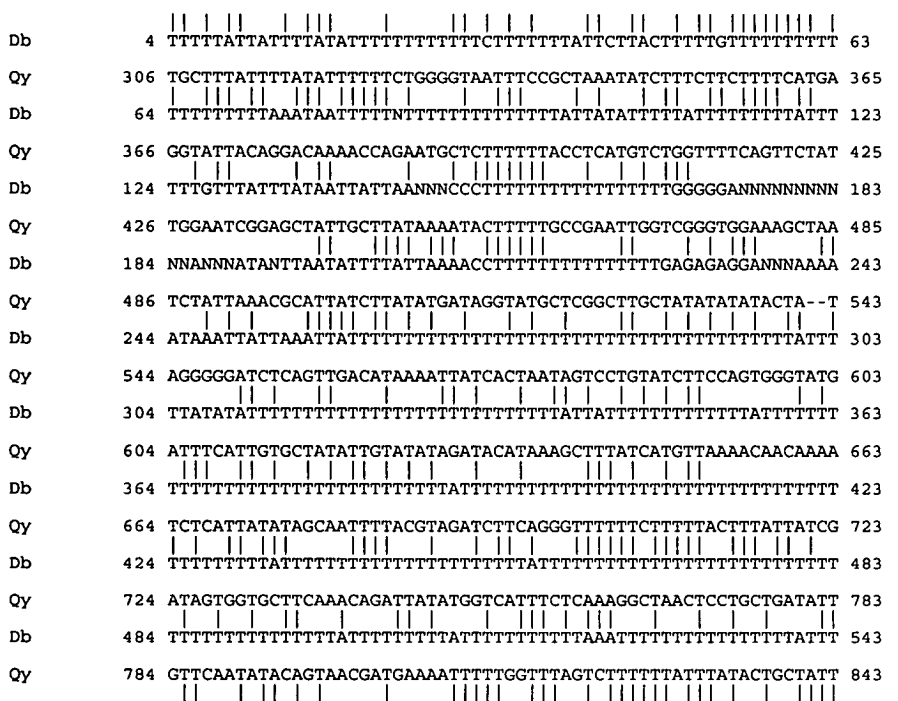
Qy       1091 CAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTACCACTACAAG 1150
Db        285 TTATTATTTAATTTAATATTATTTATTTAAATAAAAAAAAAAAAAAAATTAAAAAAAAAA 226

Qy       1151 CAATAATTGGTGGAATAGCACAA 1173
Db        225 AAAAAAAAANAAAAAAAAAAAA 203


RESULT 10
CL082000
LOCUS       CL082000                1981 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-165P18_RM4.1 CH216 Xenopus tropicalis genomic clone
            CH216-165P18, genomic survey sequence.
ACCESSION   CL082000
VERSION     CL082000.1  GI:40537913
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1981)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM4 ctcaagggcatcggtcgagc
            Class: BAC ends
            High quality sequence start: 265
            High quality sequence stop: 497.
FEATURES             Location/Qualifiers
     source          1. .1981
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-165P18"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             6.5%;  Score 83.4;  DB 10;  Length 1981;
  Best Local Similarity   43.4%;  Pred. No. 9.7e-06;
  Matches  372;  Conservative    0;  Mismatches  483;  Indels    2;  Gaps    1;

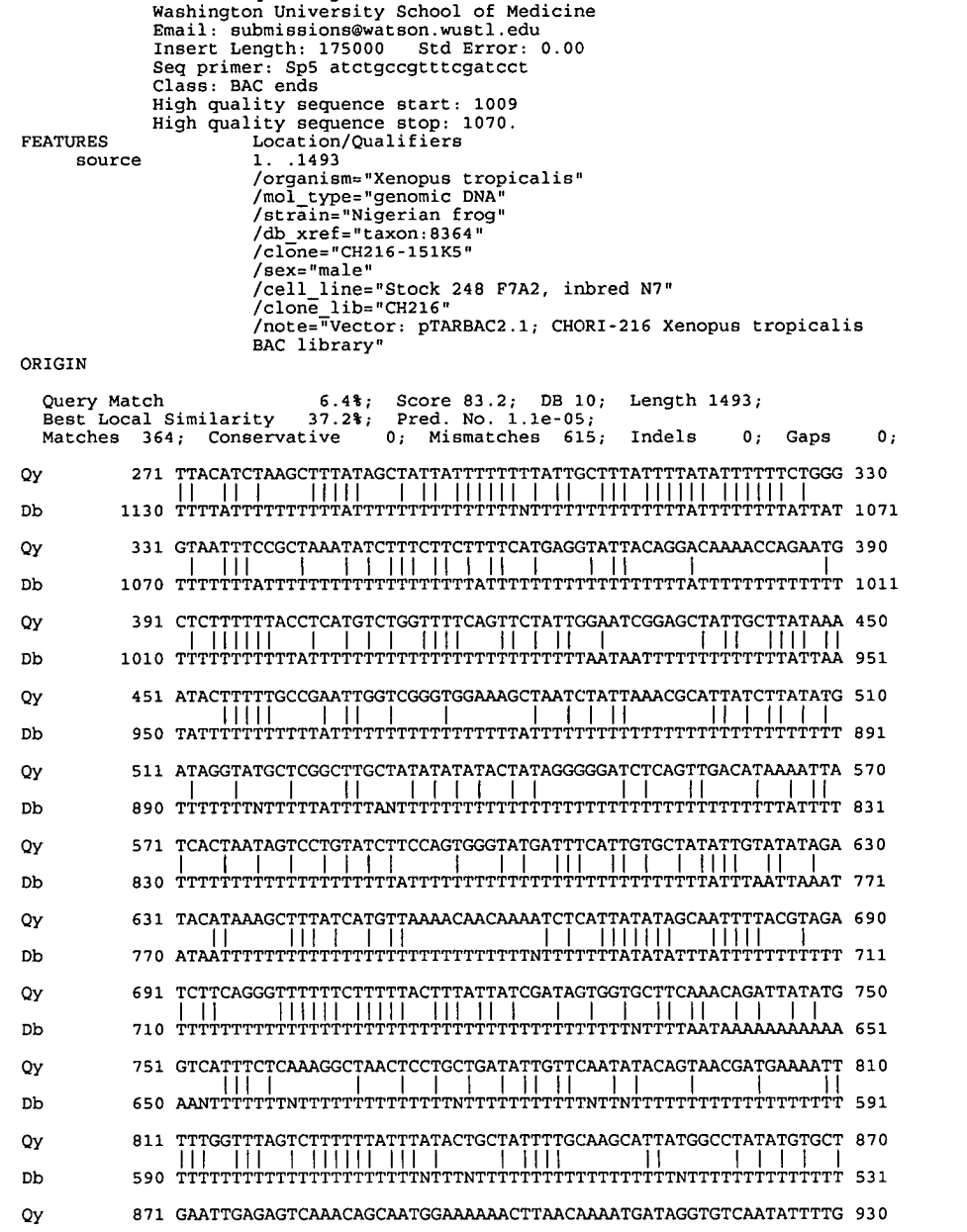
Qy        246 TTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTAT 305
```

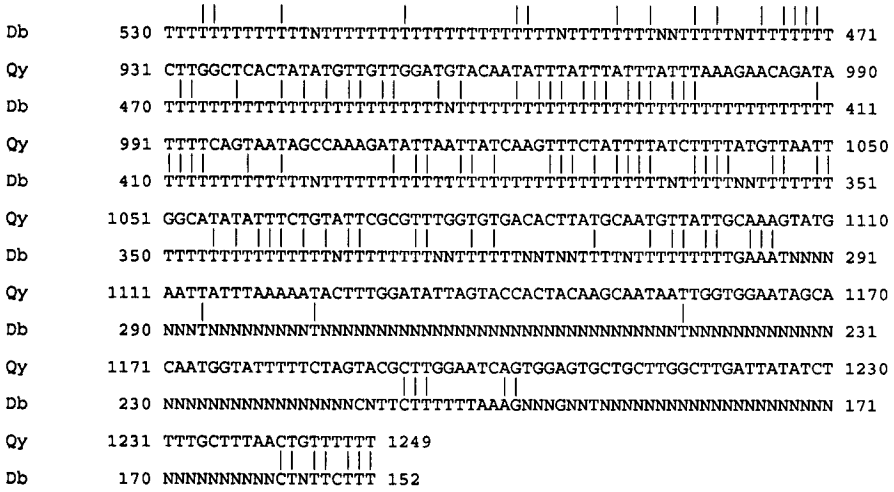


RESULT 11
CL078589/c
LOCUS CL078589 1493 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-151K5_Sp5.1 CH216 Xenopus tropicalis genomic clone CH216-151K5, genomic survey sequence.
ACCESSION CL078589
VERSION CL078589.1 GI:40534502
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 1493)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp5 atctgccgtttcgatcct
Class: BAC ends
High quality sequence start: 1009
High quality sequence stop: 1070.
FEATURES Location/Qualifiers
source 1. .1493
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-151K5"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"
ORIGIN

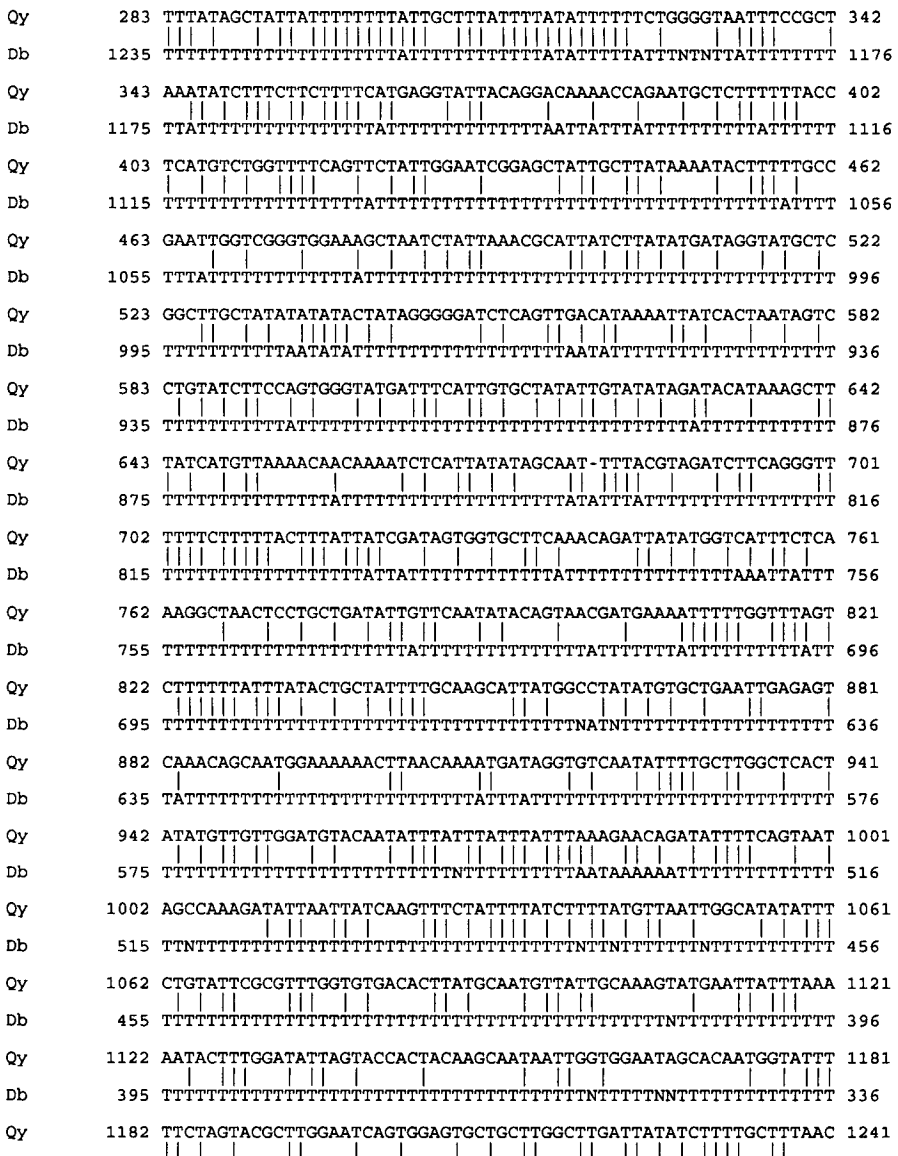
Query Match 6.4%; Score 83.2; DB 10; Length 1493;
Best Local Similarity 37.2%; Pred. No. 1.1e-05;
Matches 364; Conservative 0; Mismatches 615; Indels 0; Gaps 0;





RESULT 12
CL082569/c
LOCUS CL082569 1632 bp DNA linear GSS 31-DEC-2003
DEFINITION CH216-167P5_Sp6.1 CH216 Xenopus tropicalis genomic clone
CH216-167P5, genomic survey sequence.
ACCESSION CL082569
VERSION CL082569.1 GI:40538482
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1632)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 1062
High quality sequence stop: 1114.
FEATURES Location/Qualifiers
source 1. .1632
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-167P5"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"
ORIGIN

Query Match 6.3%; Score 81.2; DB 10; Length 1632;
Best Local Similarity 42.6%; Pred. No. 2.6e-05;
Matches 412; Conservative 0; Mismatches 555; Indels 1; Gaps 1;



RESULT 13
CL461494/c
LOCUS CL461494 1042 bp DNA linear GSS 31-MAR-2004
DEFINITION SAIL_1148_D03.v1 SAIL Collection Arabidopsis thaliana genomic clone

```
            SAIL_1148_D03.v1, genomic survey sequence.
ACCESSION   CL461494
VERSION     CL461494.1  GI:45864399
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1042)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
   PUBMED   12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS842250; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .1042
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_1148_D03.v1"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN

  Query Match             6.2%;  Score 80.8;  DB 10;  Length 1042;
  Best Local Similarity   41.8%;  Pred. No. 3e-05;
  Matches  361;  Conservative    0;  Mismatches  502;  Indels    0;  Gaps    0;

Qy        296 TTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTT 355
              |||||||| ||  ||| || | |  |  || |    |  |||    | | | | ||| ||
Db       1041 TTTTTTTTGTTTTTTTTTTATTTTATGATTTTTTTTTTTTTTTATTATTATTTTATTTNTT 982

Qy        356 CTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTT 415
              ||||  | |  ||||          |       |  | |||  ||  |  | | |   ||
Db        981 TTTTTTGTTATATATTTTTTTTATTATTTTTTTTATTTTTTAATATATTTTTTATTTATT 922

Qy        416 TCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAAATACTTTTTGCCGAATTGGTCGGGT 475
                || | ||   ||      | |   |  ||   |  |||||     |||  |    |
Db        921 ATNTTTTTTTTTTTATATTTTTTTTCAGTATTATTTTTTTTTTTTTATTATTTTAAAAT 862

Qy        476 GGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATA 535
                  | | |||||     ||  | || |||  |    ||  |    ||  | |||
Db        861 TTTTTTTTTTTTTATTATTTTTATATTATTTTATATTTTTAATATTATNATTTTTTTATT 802

Qy        536 TATACTATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAG 595
                || |||       | | | ||   | |   || | | | |||     | ||| |
Db        801 ATTATTATTTTTTTTTTTTTAATTTTTAAATTTTTTTTTATTTATATNTTTTTATTTNTATT 742

Qy        596 TGGGTATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAA 655
              |   | |  |||  || |    | || ||| || ||   || |  ||||| |  ||  |
Db        741 TTTTTTTTTTTTTTTTATTTATTTTTATATTTATATTTTTATATTTTTATTAATTTTTAT 682

Qy        656 CAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTT 715
                  | | |  || |||    |   |   |   | ||     |||||  |||||   |
Db        681 TTTATATTTTTATTTTTATTATTAAAATTTATTTTTTTTTTTTTATTTTATTTTTTTTAT 622

Qy        716 TATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTG 775
```

```
                ||||||   |  |  |     | | | | ||     |   || | |||   |   |  |
Db        621 TATTATATTTTTTTTTTTAATATATATAATAATATTTATATTATTAAATTATTTTGTTT 562

Qy        776 CTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATA 835
                | |  || |||        |      |      |||   |||  | || |||||| ||
Db        561 TATAATTATTTTAATTNTATTTTAATTATTTTATTTATTTTTTTTATTATTTATTTTTA 502

Qy        836 CTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGA 895
                |  ||| ||        || |    ||||| |  |   ||      |           |
Db        501 TTTTTATATTTATTTTTTTTTTATTTATATATTTTATTTTTTNTNTTTTTTTTATTATTA 442

Qy        896 AAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGAT 955
                || | |           |||  | | | | ||||   |   |   | | | || ||   |
Db        441 TAATAATATTATTTTATATATTTTTTATTTTTTTATATATATATTTTTTTTTTTTTTATT 382

Qy        956 GTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTA 1015
                |    ||||| |||| |   |||          | |||  |  |      |  ||||
Db        381 TTTTTTTATTTTTTTAATATATTATTTTTATTTTTTTTATAAAATATATTTTATTTATTT 322

Qy       1016 ATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTT 1075
                | |    ||| ||||||     || | ||||      |||         ||||    |||
Db        321 TATTTTTTTTTTATATTTTTATAATTTTTTTAAANNATATANNNNNTATTATTTTTATTT 262

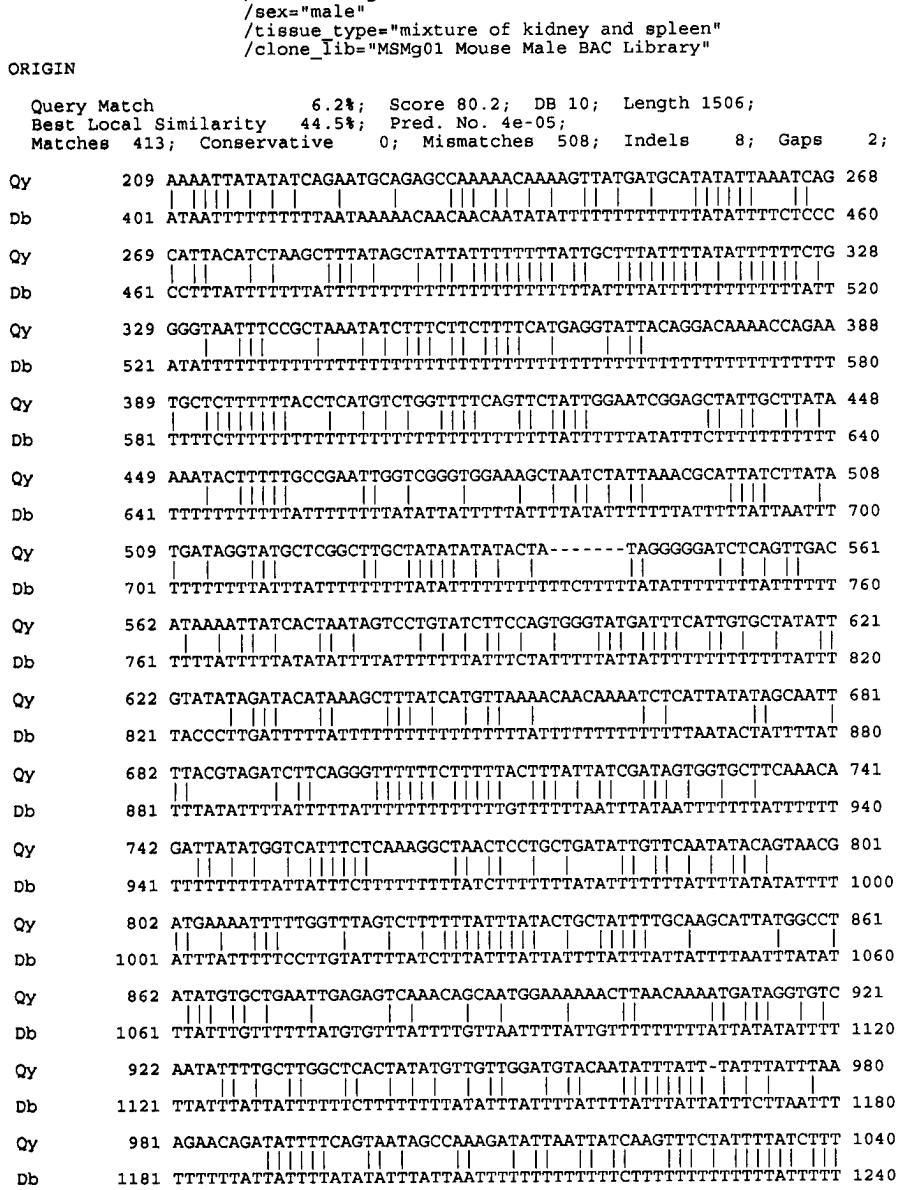
Qy       1076 GGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATAT 1135
                |     ||    ||   |||       | |   || |||      |  |||   | |
Db        261 TTNTTNNTTTTTTAAAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTNT 202

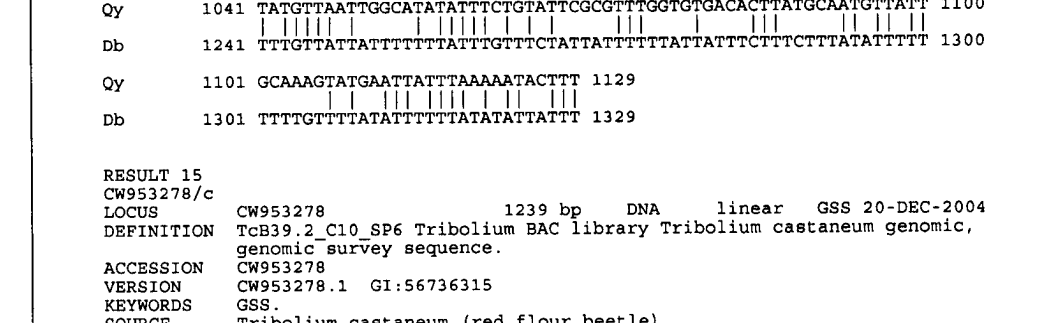
Qy       1136 TAGTACCACTACAAGCAATAATT 1158
                                 |  ||
Db        201 NTNNNNNNNNNNNNNNTTTTTTT 179
```

```
RESULT 14
AG278469
LOCUS       AG278469               1506 bp    DNA     linear   GSS 18-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-048M18.T7, genomic survey
            sequence.
ACCESSION   AG278469
VERSION     AG278469.1  GI:47851346
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1506)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
```

```
          R.Site 1   : EcoRI
          R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1506
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-048M18.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
```

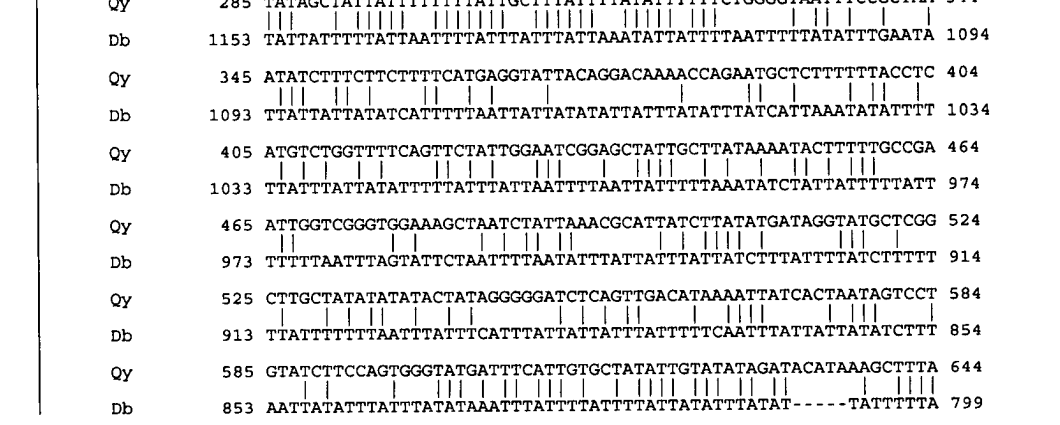
Query Match 6.2%; Score 80.2; DB 10; Length 1506;
Best Local Similarity 44.5%; Pred. No. 4e-05;
Matches 413; Conservative 0; Mismatches 508; Indels 8; Gaps 2;





```
RESULT 15
CW953278/c
LOCUS       CW953278               1239 bp    DNA     linear   GSS 20-DEC-2004
DEFINITION  TcB39.2_C10_SP6 Tribolium BAC library Tribolium castaneum genomic,
            genomic survey sequence.
ACCESSION   CW953278
VERSION     CW953278.1  GI:56736315
KEYWORDS    GSS.
SOURCE      Tribolium castaneum (red flour beetle)
  ORGANISM  Tribolium castaneum
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
            Tenebrionidae; Tribolium.
REFERENCE   1  (bases 1 to 1239)
  AUTHORS   Savard,J. and Tautz,D.
  TITLE     Tribolium castaneum BAC-ends sequencing project
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Savard, J.
            Abteilung fur Evolutionsgenetik, AG Tautz
            Institut fur Genetik, Universitat zu Koln
            Weyertal 121, 50931 Koln, Germany
            Tel: 49 221 470 6911
            Fax: 49 221 470 5975
            Email: savard@uni-koeln.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1239
                     /organism="Tribolium castaneum"
                     /mol_type="genomic DNA"
                     /strain="GA-2"
                     /db_xref="taxon:7070"
                     /clone_lib="Tribolium BAC library"
                     /note="Vector: pBACe3.6; Site_1: EcoR1; Site_2: EcoR1;
                     Library constructed by Exelixis Inc."
ORIGIN
```

Query Match 6.2%; Score 79.6; DB 10; Length 1239;
Best Local Similarity 45.1%; Pred. No. 5.2e-05;
Matches 384; Conservative 0; Mismatches 459; Indels 9; Gaps 2;



```
Qy        645 TCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTT 704
              | || |  |    |  ||  | |  |||  ||  |  ||     ||     || |   |||||
Db        798 TTATATATATTTATTAATTTATTTTTAATTATTATATTATTTTATTATTTTATTTTTTTT 739

Qy        705 TCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAG 764
              |  |||   ||| | ||   || |  |  ||         |  |||  | || | |
Db        738 TAATTTATATTTGTGATTATTATTTATTATTATTTTGATTATTATATTAATATTTTTTTT 679

Qy        765 GCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTT 824
                ||  |  |    ||||  | | | ||  |||          | |   | | || |
Db        678 TGTATATTAT----TTATTTATAATTTTAATGTAGTTTATTTTTATGTCATTATTGTTTA 623

Qy        825 TTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAA 884
              |||| | | || || ||| ||  |     | |    ||| | |  |||   |    |
Db        622 TTTTTTATTTATTGTTATATTATATTTTATTTTTATTATTTTTTTTGATCAGTTTATATG 563

Qy        885 ACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATA 944
              |     ||       | |||| |  ||  |       ||| || |  ||     | | ||
Db        562 ATTTTCATATTTTTTATTTAAAATTATTTTTATATATAATTTTATTATTATTAAAATTTA 503

Qy        945 TGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGC 1004
              | || ||   | |   || | | ||||||| |||||   |    |||| |   || |
Db        502 TTTTCTTTATTATTATATTTATCTTTATTTTTTTAATTTATTATTATTATTTATATTTAT 443

Qy       1005 CAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTG 1064
                | || || |||||  |   | | ||||||  | |  | ||| ||    | |||||  |
Db        442 TTTAAATTTTTATTATTTATAATTTTTTTTATTTATATTATTATTTTCATTTTATTTTTT 383

Qy       1065 TATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAAT 1124
              |||     |||  |    | ||||  | | |  ||  | | | |  | |||||   | |
Db        382 TATATATATTTATTTATTTATTTATATATTTTATTTTAATATTTTTCAATATTTTTTATT 323

Qy       1125 ACTTTGGATATT 1136
              |  ||   || |
Db        322 ATATTATTTAAT 311
```

Search completed: May 12, 2006, 03:30:37
Job time : 4047.56 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 19:15:19 ; Search time 176.843 Seconds
(without alignments)
12996.800 Million cell updates/sec

Title: US-09-423-093B-4_COPY_12762_14054
Perfect score: 1293
Sequence: 1 gtgaaagttcaattgttaaa.....ttaaggcaaataagggataa 1293

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	67	5.2	1039	3	US-09-902-540-1280	Sequence 1280, Ap
	2	61.2	4.7	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	3	60.6	4.7	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	4	60.2	4.7	612	3	US-09-902-540-1357	Sequence 1357, Ap
c	5	59.6	4.6	187169	3	US-09-949-016-12776	Sequence 12776, A
c	6	59.6	4.6	191569	3	US-09-949-016-15940	Sequence 15940, A
c	7	59.4	4.6	700	3	US-09-735-271-1038	Sequence 1038, Ap
	8	59.2	4.6	18773	3	US-09-949-016-14164	Sequence 14164, A
c	9	58.8	4.5	18773	3	US-09-949-016-14164	Sequence 14164, A
c	10	58.6	4.5	1132	3	US-09-533-559-5452	Sequence 5452, Ap
c	11	57.8	4.5	205044	3	US-09-949-016-15851	Sequence 15851, A
c	12	57.8	4.5	205044	3	US-09-949-016-15852	Sequence 15852, A
c	13	57.8	4.5	205044	3	US-09-949-016-15853	Sequence 15853, A
c	14	57.8	4.5	223471	3	US-09-949-016-12387	Sequence 12387, A
c	15	57.8	4.5	223471	3	US-09-949-016-12724	Sequence 12724, A
c	16	57.8	4.5	223471	3	US-09-949-016-12725	Sequence 12725, A
c	17	56.6	4.4	129415	3	US-09-949-016-16997	Sequence 16997, A
	18	56.4	4.4	205044	3	US-09-949-016-15851	Sequence 15851, A
	19	56.4	4.4	205044	3	US-09-949-016-15852	Sequence 15852, A
	20	56.4	4.4	205044	3	US-09-949-016-15853	Sequence 15853, A
	21	56.4	4.4	223471	3	US-09-949-016-12387	Sequence 12387, A
	22	56.4	4.4	223471	3	US-09-949-016-12724	Sequence 12724, A
	23	56.4	4.4	223471	3	US-09-949-016-12725	Sequence 12725, A
c	24	55.4	4.3	601	3	US-09-949-016-30531	Sequence 30531, A
c	25	55.4	4.3	601	3	US-09-949-016-37150	Sequence 37150, A
c	26	55.4	4.3	601	3	US-09-949-016-37164	Sequence 37164, A
c	27	55.4	4.3	601	3	US-09-949-016-145868	Sequence 145868,
c	28	55.4	4.3	601	3	US-09-949-016-146136	Sequence 146136,
c	29	55.4	4.3	601	3	US-09-949-016-146404	Sequence 146404,
	30	54.8	4.2	700	3	US-09-735-271-1038	Sequence 1038, Ap
c	31	54.6	4.2	601	3	US-09-949-016-54020	Sequence 54020, A
c	32	54.6	4.2	601	3	US-09-949-016-108654	Sequence 108654,
	33	54.4	4.2	5852	2	US-07-867-106-2	Sequence 2, Appli
	34	53.8	4.2	96922	3	US-09-949-016-17061	Sequence 17061, A
	35	53.8	4.2	129415	3	US-09-949-016-16997	Sequence 16997, A
	36	53.4	4.1	134987	3	US-09-949-016-15348	Sequence 15348, A
	37	53.4	4.1	134987	3	US-09-949-016-15349	Sequence 15349, A
	38	53.4	4.1	134987	3	US-09-949-016-15350	Sequence 15350, A
	39	53.4	4.1	134987	3	US-09-949-016-15507	Sequence 15507, A
	40	53.4	4.1	134987	3	US-09-949-016-15508	Sequence 15508, A
	41	53.4	4.1	134987	3	US-09-949-016-15509	Sequence 15509, A
	42	53.4	4.1	640681	3	US-09-790-988-1	Sequence 1, Appli
c	43	53	4.1	601	3	US-09-949-016-54021	Sequence 54021, A
c	44	53	4.1	601	3	US-09-949-016-54024	Sequence 54024, A
c	45	53	4.1	601	3	US-09-949-016-108655	Sequence 108655,

ALIGNMENTS

RESULT 1
US-09-902-540-1280/c
; Sequence 1280, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1280
; LENGTH: 1039
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1039)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1280

Query Match 5.2%; Score 67; DB 3; Length 1039;
Best Local Similarity 43.1%; Pred. No. 0.00018;
Matches 430; Conservative 0; Mismatches 557; Indels 10; Gaps 2;

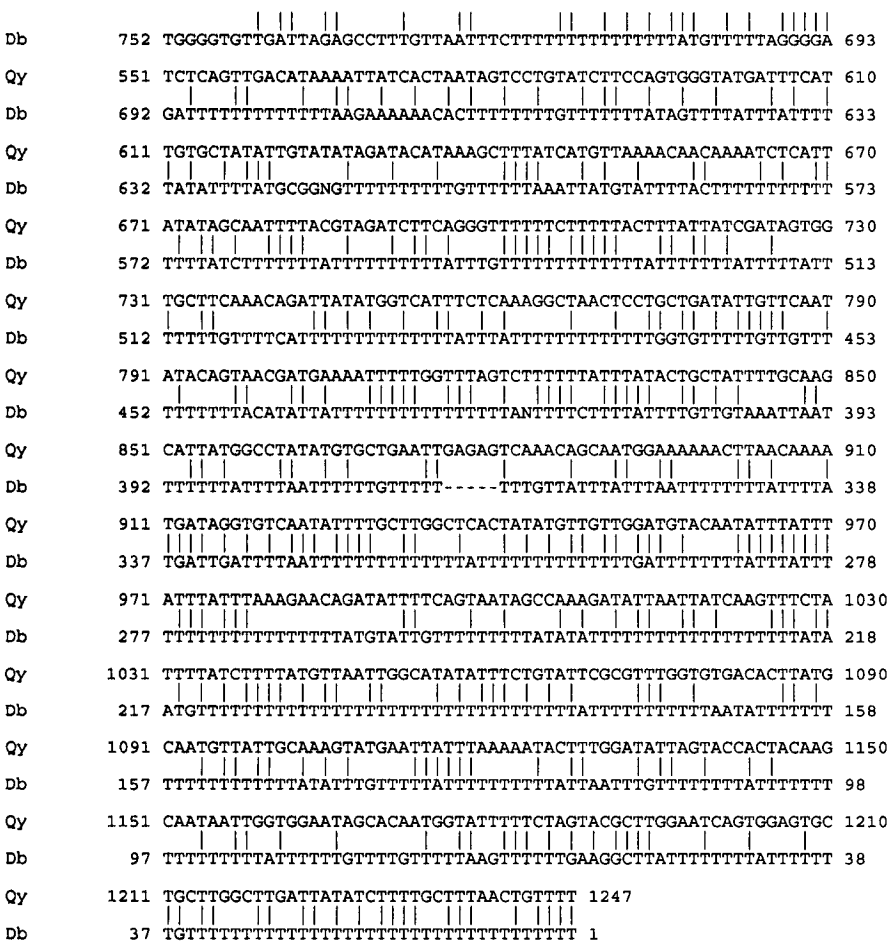
```
Qy        256 TATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTT 315
              | | ||  || ||  |||  | |    ||| |   | |    |||| | ||  ||| |||
Db        992 TTTTTTTTATTAGTTTTATGTTTTTTTTTTTTTTTTTTTATGGTTTTATTTTTTTTTTTT 933

Qy        316 TATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAG 375
              |  || |||| |    |  |||   ||      | |||  | ||||  |  | | ||
Db        932 TGGATATTTTGTAATTTTTTTTTTTCTTTTATTTTTTTATTTTTTATTTGGTTTTTTATT 873

Qy        376 GACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGA 435
                  |  | |  | |||||||  |   ||   |||| ||| | ||    |
Db        872 TTTTTTAAGTTATTTTTGTTTTTTATTTTCCCTCATTTTTTTAGTGTTTTTAATTTTTTT 813

Qy        436 GCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAAC 495
                | ||    |       |||||      ||  |  | |   | | |  | | ||| |
Db        812 TTTTTTTAGAAGGTTTATATTTTTATATTTTTTTTTTGTTTTTAGGATTTTTTGTTATAT 753

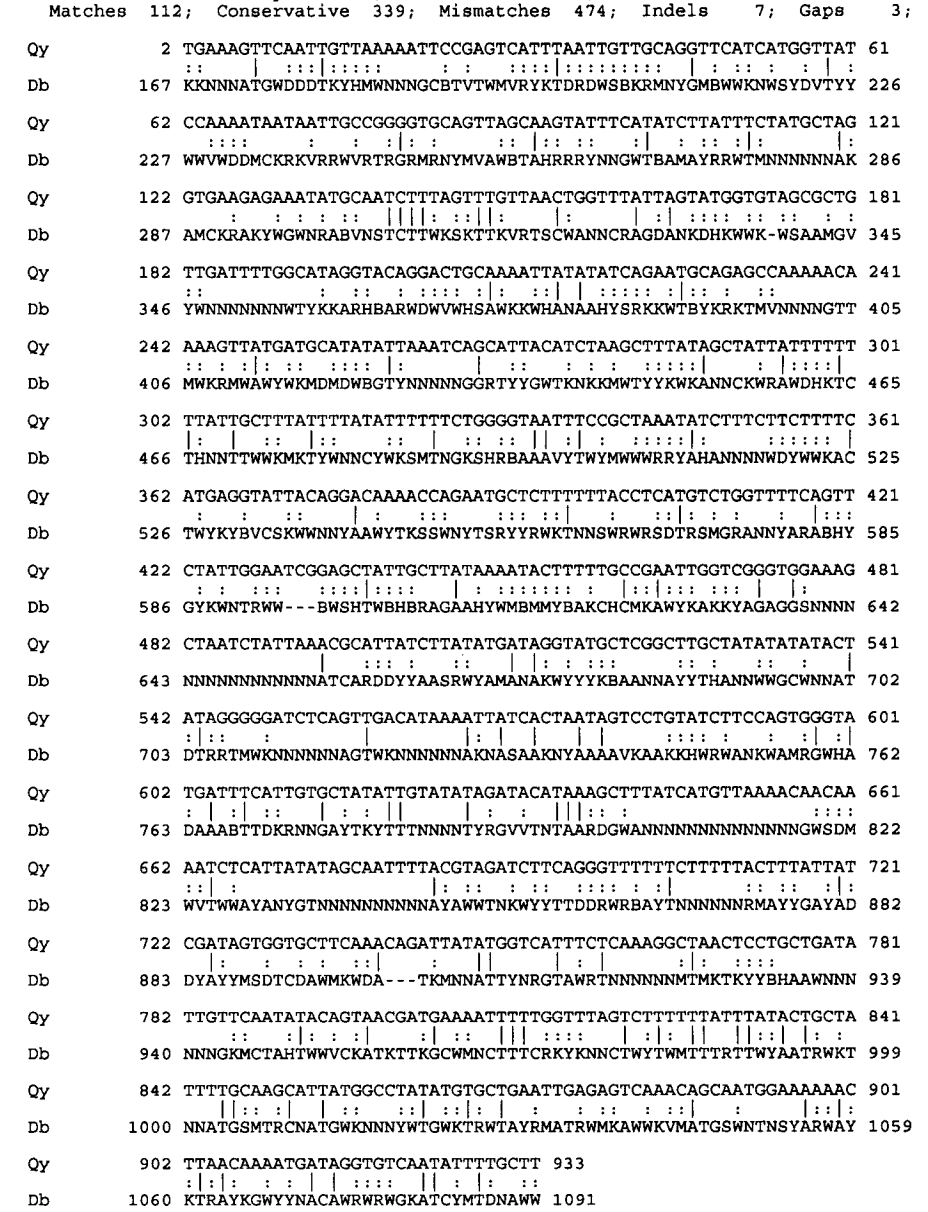
Qy        496 GCATTATCTTATATGA-----TAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGA 550
```



RESULT 2
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)

; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 4.7%; Score 61.2; DB 3; Length 1141;
Best Local Similarity 12.0%; Pred. No. 0.0026;
Matches 112; Conservative 339; Mismatches 474; Indels 7; Gaps 3;



RESULT 3

```
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             4.7%;  Score 60.6;  DB 3;  Length 1141;
  Best Local Similarity   11.9%;  Pred. No. 0.0034;
  Matches  118;  Conservative  358;  Mismatches  512;  Indels    4;  Gaps    3;

Qy         42 TGCAGGTTCATCATGGTTATCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTC 101
Db       1058 TWYTRSNANWSCATKBMWWTMKWYATKYRTAWYAMWCAWRNNNMWCATNGYAKSCATNNA 999

Qy        102 ATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTT 161
Db        998 MWYATTRWAAYAAAKWARWAGNNMRMYGAAAGNKWGCMAAMATMGBWWADTAGKMCNNNN 939

Qy        162 ATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTA-CAGGACTGCAAAATTATATAT 220
Db        938 NNWTTDVRRMAMKAKNNNNNNAYWTACYNRAATNNKMATHWMKWTHGAHSKRRTRHHTRT 879

Qy        221 CAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTACATCTAA 280
Db        878 CRRTKYNNNNNNARTVYWYHHAARRWMNAWWTRTNNNNNNNNNNACRNTRTWWABWKHSW 819

Qy        281 GCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCG 340
Db        818 CNNNNNNNNNNNNNTWCHYTTANABBCYRANNNNAAARMARTCNNYMHAAVTTTHTDWC 759

Qy        341 CTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTA 400
Db        758 YKTWMNTWYWDMMTTMBTTTTRNMTTSTNMTNNNNNNMWACTNNNNNNMWKAYYAHATNN 699

Qy        401 CCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTG 460
Db        698 WGCWWNNTDARRTNNTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNNNNNNNNNNNNNNN 639

Qy        461 CCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTATGC 520
Db        638 SCCTCTRMMTMRWTMKGDGMTVRKKVKWRDTTCTYVDVWADSWVWWYANWMRCRDVTYTR 579

Qy        521 TCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAG 580
Db        578 NNTYCKSYAHSYWYWSNNAMWYRRYSARNWSSMARWTTRNNWWMSGBVRMRWAGTMWWRH 519

Qy        581 TCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACATAAAGC 640
Db        518 WNNNNTDTRYYWWWKRWARBTTTVYD-SMCNAKSMWRGNNWRAMKMWWAANNDAGAMDHW 460

Qy        641 TTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGT 700
Db        459 TYWMGNNTMWMRRAWKMMNMAWCRRAYCCNNNNNRACVWHKHKMWRWTWKYMWKAACNNN 400

Qy        701 TTTTTCTTTTTACT--TTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTC 758
Db        399 NBKAMYMRVAWMMYSRDTTNTDWMMWTSDWBWHWYTVDYTMMRAWNNNNNNNWRBCKTTS 340

Qy        759 TCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTT 818
Db        339 WMWWMDHMNTHCTYGNNTWGSAYBMAAMSMWAAGASNBVTYNWCWRMTYMGKTMTNNNNN 280

Qy        819 AGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAG 878
Db        279 NKAWYYRTKTVAWCNNRYYYDTAVWTBKRNYKYCYAYBWYYBMYMGKHHWBWWRRABHRS 220

Qy        879 AGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTC 938
Db        219 WNMWWVKCRNKYMVSWHYHAMRYBKWABAVGCNNNWKDRMAHHHWCATNNNMMWWWYAYM 160

Qy        939 ACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGT 998
Db        159 HHMHKKGKAAWTNNKTABRDDHBAHVKTYWYWRYDYWCAMCWMNAKAKVRTAMKHMWYYT 100

Qy        999 AATAGCCAAAGATATTAATTATCAAGTTTCTA 1030
Db         99 DRYVSANNTGVRWMMMRWCMWWYSMNNRWYYR 68


RESULT 4
US-09-902-540-1357/c
; Sequence 1357, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
;  APPLICANT: Goldman, Barry S.
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15849)B
;  CURRENT APPLICATION NUMBER: US/09/902,540
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 60/217,883
;  PRIOR FILING DATE: 2000-07-10
;  NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1357
;   LENGTH: 612
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(612)
;   OTHER INFORMATION: unsure at all n locations
US-09-902-540-1357

  Query Match             4.7%;  Score 60.2;  DB 3;  Length 612;
  Best Local Similarity   44.0%;  Pred. No. 0.0039;
  Matches  248;  Conservative    0;  Mismatches  315;  Indels    0;  Gaps    0;

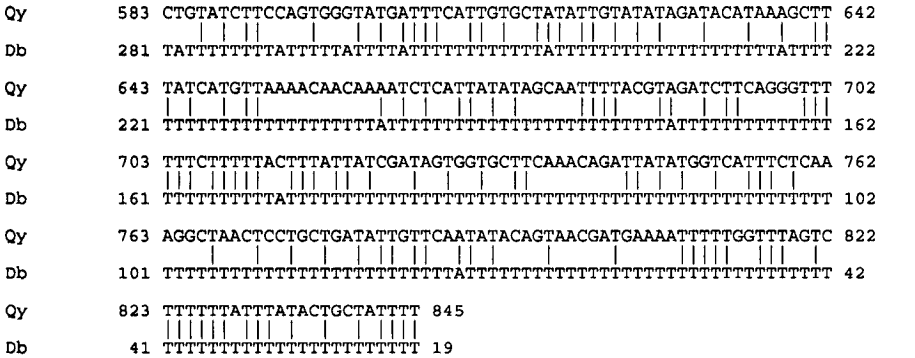
Qy        283 TTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCT 342
Db        581 TTTTTATTTTTTTTTTTTTTTTTTTTTAATTTTTTTTTTTTTTTTTTTTTTTTTTTTT 522

Qy        343 AAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACC 402
Db        521 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTA 462

Qy        403 TCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCC 462
Db        461 TTATTTTTTAATTTTTTTTTAATTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTT 402

Qy        463 GAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTC 522
Db        401 TTTTATTTTTATATTTTATTTTTTTTTTTTTATTTTTTTTTTTATTTATTTTTTTTTATT 342

Qy        523 GGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTC 582
Db        341 TTTTTANTTTTTTTTTTTTTTTATTTTANTTTTTTTTTTTTTTAATTTTTTTTTAATATTT 282
```



RESULT 5
US-09-949-016-12776/c
; Sequence 12776, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12776
; LENGTH: 187169
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(187169)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12776

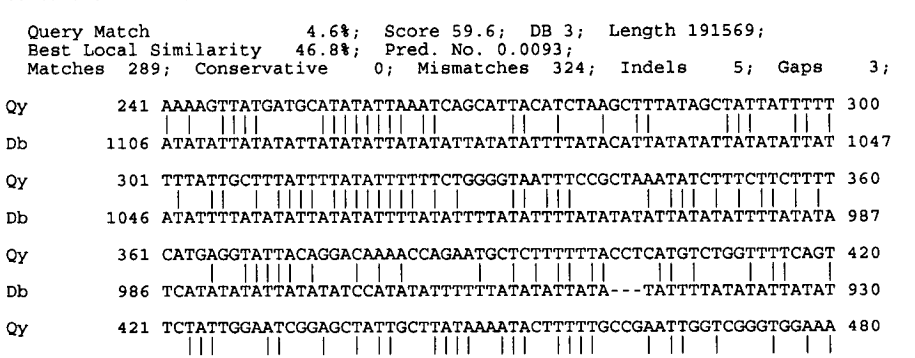
Query Match 4.6%; Score 59.6; DB 3; Length 187169;
Best Local Similarity 46.8%; Pred. No. 0.0093;
Matches 289; Conservative 0; Mismatches 324; Indels 5; Gaps 3;





RESULT 6
US-09-949-016-15940/c
; Sequence 15940, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15940
; LENGTH: 191569
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(191569)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15940

Query Match 4.6%; Score 59.6; DB 3; Length 191569;
Best Local Similarity 46.8%; Pred. No. 0.0093;
Matches 289; Conservative 0; Mismatches 324; Indels 5; Gaps 3;



```
Db        929 ATTATATATATTATATATTTTATATATATTATATATTTTATATATTTTATATATTTTATA 870

Qy        481 GCTAATCTATTAAACGCAT-TATCTTATATGATAGGTATGCTCGGCTTGCTATATATATA 539

Db        869 TATATTTTATATATTTTATATATTATATATATTTTATATATTATATATTTTATATATTTT 810

Qy        540 CTATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGG 599

Db        809 ATATATTTTATATATTATATATAATATATTTTATATTTTATATATTATATATTTTA 750

Qy        600 TATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAAC 659

Db        749 TATTTTTTATATATATTATATATTTTATATTTTATATATTATATATTTATATATATT 690

Qy        660 AAAATCTCATTATATAGCA-ATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTAT 718

Db        689 TTATGTATATTATATATTATATATATTTTATATATTATATATTATATATTTTATATAT 630

Qy        719 TATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTG 778

Db        629 TATATATTATATATATTTATATATATTATATAGTATATATATTTATATATACTATATAAT 570

Qy        779 ATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTG 838

Db        569 ATATATTTTATATATATTATATATTATATATATTTTATATATTATATATATTTTATATAT 510

Qy        839 CTATTTTGCAAGCATTAT 856

Db        509 ATTTTTTATATATTTTAT 492


RESULT 7
US-09-735-271-1038/c
; Sequence 1038, Application US/09735271
; Patent No. 6869762
; GENERAL INFORMATION:
;  APPLICANT: Daly, Mark J.
;  APPLICANT:  Hudson, Thomas J.
;  APPLICANT:  Lander, Eric S.
;  APPLICANT:  Rioux, John
;  APPLICANT:  Siminovitch, Kathy
;  TITLE OF INVENTION: IBD-RELATED POLYMORPHISMS
;  FILE REFERENCE: 2825.1025-002
;  CURRENT APPLICATION NUMBER: US/09/735,271
;  CURRENT FILING DATE:  2000-12-11
;  PRIOR APPLICATION NUMBER: US 60/170,257
;  PRIOR FILING DATE: 1999-12-10
;  PRIOR APPLICATION NUMBER: US 60/196,046
;  PRIOR FILING DATE: 2000-04-10
;  NUMBER OF SEQ ID NOS: 2058
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1038
;   LENGTH: 700
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(700)
;   OTHER INFORMATION: n = A,T,C or G
US-09-735-271-1038

  Query Match             4.6%;  Score 59.4;  DB 3;  Length 700;
  Best Local Similarity   45.4%;  Pred. No. 0.0057;
  Matches  251;  Conservative    0;  Mismatches  301;  Indels    1;  Gaps    1;

Qy        485 ATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTATA 544

Db        590 ATATTATATATTTAAAATATATATTAATATATATTATATATATTTAAAATATATAATATA 531

Qy        545 GGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGA 604

Db        530 TATTAAACATATTTAAAATATATATTTAAAAATATAAATTATATATTATATATAATATAT 471
```

```
Qy        605 TTTCATTGTGCTATATTGTATAT-AGATACATAAAGCTTTATCATGTTAAAACAACAAAA 663

Db        470 ATTATATATTATATATAATATATAATATATATAATATATAATATATATTATATATAAT 411

Qy        664 TCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCG 723

Db        410 ATATATTATATATTATATATTTATATATTTATATATTGTATATTTATATATTGTATATAT 351

Qy        724 ATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATT 783

Db        350 TTATATATTATATATTTATATATTTTTATATATTATATATATTTATATATATTATATATA 291

Qy        784 GTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATT 843

Db        290 TTTTTATATATATTTTTATATATTATATATTTATATATATTTTTATATATTTTTATATAT 231

Qy        844 TTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTT 903

Db        230 TTTATATATTTATATATATTTTTATATATTATATATATTTATATATATTTTTATATATTA 171

Qy        904 AACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATA 963

Db        170 TATATATTTATATATATTTATATATTATATATATTTTTATATATTTTTATATATTACATA 111

Qy        964 TTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAA 1023

Db        110 TATATTTATATATATTTTTATATATTATATATATTTATATATATATATTTTAGTGCCCAA 51

Qy       1024 GTTTCTATTTTAT 1036

Db         50 TAGGTAGTTTTAT 38


RESULT 8
US-09-949-016-14164
; Sequence 14164, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14164
;   LENGTH: 18773
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-14164

  Query Match             4.6%;  Score 59.2;  DB 3;  Length 18773;
  Best Local Similarity   43.2%;  Pred. No. 0.0088;
  Matches  393;  Conservative    0;  Mismatches  508;  Indels    9;  Gaps    2;

Qy        213 TTATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATT 272

Db      17335 TTATATATAATATAATATATATTTTATATATAATTATTAATTATATAATTAATATATAAT 17394

Qy        273 ACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGT 332

Db      17395 ATATATTTTATACATAATTATTAATTATATATAATTAATATATAATATATCTTATACATA 17454

Qy        333 AATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCT 392

Db      17455 ATTATCAATTATATATAATTAATATATAATATATATTTTATATATAATATAATATATTTT 17514
```

```
Qy        393 CTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAAT 452
Db      17515 ATATATTATATATTATATATAATATAATATGTATTCTATATATTATATTATATAAAATAT 17574

Qy        453 ACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGAT 512
Db      17575 AGTATATAATATATAATATATAATATATAATACAATATATAATATAATATATAAT 17634

Qy        513 AGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATC 572
Db      17635 ATAAAAATATATAATATAATATATAATATATAATATAATAT--------ATAAATGTATA 17686

Qy        573 ACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATA 632
Db      17687 ATATATAATATAATATATAATGTATAATATATAATATATAATGTATAATATATAATATAT 17746

Qy        633 CATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATC 692
Db      17747 ATTATATATTATATATATATTATATGTAATATATAATATATAATATATAATATATAATAT 17806

Qy        693 TTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGT 752
Db      17807 ATAATATATAATATATATTATATTATATATAATATATAATATATATAATATATAATATTA 17866

Qy        753 CATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTT 812
Db      17867 TATATAATATATAATATATATTATATTATATATGATATATATTTTATATATTATATTATA 17926

Qy        813 TGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGA 872
Db      17927 TATGATATATATTTTATATATATATGAAATTTATATATACTATTATATATAATATTATAT 17986

Qy        873 ATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCA-ATATTTTGC 931
Db      17987 ATTTTATATATAATATATATTATATAATATTATATATATTATACATTACATATATATTAT 18046

Qy        932 TTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATAT 991
Db      18047 ATATAATATTATATATTTTATATACAATATGATATATAATATAATTTATATATTATATAT 18106

Qy        992 TTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTG 1051
Db      18107 ATTTATATATAATTATTATATAAATTATATAAATATAAATTATATATTTATATATAATTA 18166

Qy       1052 GCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGA 1111
Db      18167 TTATATAAATCATTATATAATTATTATAATTATAATATATAATATAATATAATTATTAT 18226

Qy       1112 ATTATTTAAA 1121
Db      18227 ATAATATATA 18236
```

RESULT 9
US-09-949-016-14164/c
; Sequence 14164, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 14164
; LENGTH: 18773
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14164

Query Match 4.5%; Score 58.8; DB 3; Length 18773;
Best Local Similarity 44.8%; Pred. No. 0.011;
Matches 433; Conservative 0; Mismatches 522; Indels 11; Gaps 5;

```
Qy         96 TATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTTAAC 155
Db      18105 TATATAATATATAAATTATATTATATATCATATTGTATATAAAATATATAATATTATATA 18046

Qy        156 TGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAATTA 215
Db      18045 TAATATATATGTAATGTATAATATATAATATTATATAATATATATTATATATAAAATA 17986

Qy        216 TATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTACA 275
Db      17985 TATAATATTATATATAATAGTATATATAAATT-TCATATATATATAAAATATATATCATA 17927

Qy        276 TCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAAT 335
Db      17926 TATAATATAATATATAAAATATATATCATATATAATATAATATATATTATATATTATATA 17867

Qy        336 TTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTT 395
Db      17866 TAATATTATATATTATATATATTAT----ATATTATATATAATATAATATATATTATATA 17811

Qy        396 TTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACT 455
Db      17810 TTATATATTATATATTATATATTATATATTATATATTACATATAATATATATATAATATA 17751

Qy        456 TTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGG 515
Db      17750 TAATATATATTATAT----ATTATACATTATATATTATATATTATACATTATATATTATA 17695

Qy        516 TATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACT 575
Db      17694 TTATATATTATACATTTATATATTATATTATATATTATATATTATATTATATATTTTTAT 17635

Qy        576 AATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACAT 635
Db      17634 ATTATATATTATATTATATATTGTATTATATATTATATATTATATTATATATTATATACT 17575

Qy        636 AAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTC 695
Db      17574 ATATTTTATATAATATAATATATAGAATACATATTATATTATATATAATATATAATATAT 17515

Qy        696 AGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCAT 755
Db      17514 AAAATATATTATATTATATATAAAATATATATTATATATTAATTATATATAATTGATAAT 17455

Qy        756 TTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTA-ACGATGAAAATTTTTG 814
Db      17454 TATGTATAAGATATATTATATATTAATTATATATAATTAATAATTATGTATAAAATATAT 17395

Qy        815 GTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAAT 874
Db      17394 ATTATATATTAATTATATAATTAATAATTATATATAAAATATATATTATATTATATATAA 17335

Qy        875 TGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTG 934
Db      17334 AAATATTATATATAAAATATTATATATTAAATAATATTATTTTATATGTTATATATTATA 17275

Qy        935 GCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTT 994
Db      17274 CATTATTAGATATTCTATATTATATTTTATATAATTATATA-TAATTATATACACATTTT 17216

Qy        995 CAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCA 1054
Db      17215 ATATAGAAAATATTATATATAATATTATATATAATATTTTATATATAAAATTATATATAA 17156
```

```
Qy       1055 TATATT 1060
              |||||
Db      17155 AATATT 17150
```

RESULT 10
US-09-533-559-5452/c
; Sequence 5452, Application US/09533559
; Patent No. 6902887
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Jeffrey R.Shuster
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjarke Olsen
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 5849.200-US
; CURRENT APPLICATION NUMBER: US/09/533,559
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 09/273,623
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5452
; LENGTH: 1132
; TYPE: DNA
; ORGANISM: Aspergillus oryzae
US-09-533-559-5452

Query Match 4.5%; Score 58.6; DB 3; Length 1132;
Best Local Similarity 46.1%; Pred. No. 0.0086;
Matches 269; Conservative 0; Mismatches 309; Indels 5; Gaps 2;

```
Qy        557 TTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCT 616
              || | || ||  | | | || || || |  | |   | |    ||  |       |  |
Db       1117 TTAATATTAATATTTTAGTATTATTCATAATTTTATTATTATTCATTTTCAATGCTTATT 1058

Qy        617 ATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAG 676
              || || | |||   || |   |   ||   || |   | | ||  ||| ||||||  ||
Db       1057 ATTTTATTTATTATTATACTCATAGTTTCAATTTGTTATCTACTCAATGTCATTACATTT 998

Qy        677 CAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTC 736
              ||  |||      ||  |  |   |   || | |    ||||||||   | |  |  ||
Db        997 CATATTTCTAATGAAAT--ATTTTAAATTATATAACTTTTATTATGTCAATTATTAATTT 940

Qy        737 AAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAG 796
                |  | |  ||  |  ||| | | |   ||||        |||   || |||||| |
Db        939 TTAAGAAATTAATTTTATTTTATTATAATATAACATAATAATATAAATTTTAATATATAT 880

Qy        797 TAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTAT 856
              |      | ||| | |    ||   | ||  || ||||   |     |||  || | ||||
Db        879 TTTATTTTAAATTATAATTTTATTTATTAATTCTTTAATATATATCTTTTGAAACTTTAT 820

Qy        857 GGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAG 916
              |   || || |  |  ||| | | |          ||  ||  |    |  |||
Db        819 G---TAAATTTTATTTATTTAAAATTTTATCGTTAATTTTAATATTTAATTTAATATAA 763

Qy        917 GTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTAT 976
              | |   ||||||  |    |   | ||  || |  ||    ||| ||||| ||| |||
Db        762 TTTTTTTTATTTTATTATTATATATTTAATTTTTAAAATTATTAATTTTTATCTATATAT 703

Qy        977 TTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTAT 1036
              || |    |  |  ||| |    ||   | |  || || || |    ||  ||| |  |
Db        702 ATATATTTTATTTTATTTAATATTTATTTATAATTAATATTTTTTTGTTTAATATATATT 643

Qy       1037 CTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGT 1096
              | ||||||   |   || ||| |    |     |   |       || |  ||| |
Db        642 ATATTATGTATTTCTAAATTTATATACCAAAGTAAAATCTATAAATTTTTTTTATAATAT 583

Qy       1097 TATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGT 1139

```
              ||   | | ||| | || || | | ||    ||   ||  ||
Db        582 GATATAATATTATAATTTTTTAATATATTTAATGTTAATATGT 540
```

RESULT 11
US-09-949-016-15851/c
; Sequence 15851, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15851
; LENGTH: 205044
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(205044)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15851

Query Match 4.5%; Score 57.8; DB 3; Length 205044;
Best Local Similarity 45.8%; Pred. No. 0.021;
Matches 237; Conservative 0; Mismatches 277; Indels 3; Gaps 1;

```
Qy        609 ATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCA 668
              |||    | ||   |||||| ||   || |  |||||  |   || | |   | | |
Db     201081 ATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTAT 201022

Qy        669 TTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGT 728
              ||| ||    |   |  || || ||       | ||| | ||||  |||      | |
Db     201021 ATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTT 200962

Qy        729 GGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCA 788
                || | |   |  ||||    ||||  |  |    ||  |       ||||   |
Db     200961 ATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTT 200902

Qy        789 ATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCA 848
              |||| | | |   | ||||| |    | || | || | |||||||| |   | | | |
Db     200901 TTATATATTTATATTTAAAATAT---ATATATTTTTATATATTTATATTTAAAATATATA 200845

Qy        849 AGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAA 908
                ||     | ||| |    ||| | |       |   ||  | |     |   |
Db     200844 TATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATA 200785

Qy        909 AATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTAT 968
              || |       | |||| ||  |       | ||||| |   |  || |  |||||||
Db     200784 TATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTAT 200725

Qy        969 TTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTC 1028
              | |  | |  |   | | |||||||   || |       | ||  ||    |  | |
Db     200724 ATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATT 200665

Qy       1029 TATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTA 1088
              |  | ||| | |   |   |||    ||| || | | |   |    ||   | | | | ||
Db     200664 TTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATA 200605

Qy       1089 TGCAATGTTATTGCAAAGTATGAATTATTTAAAAATA 1125
              |  ||   ||||   |  |||  ||  || ||| |||
Db     200604 TATAAATATATTTTTATATATTTATATTTAAAATATA 200568
```
```

```
RESULT 12
US-09-949-016-15852/c
; Sequence 15852, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15852
;   LENGTH: 205044
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(205044)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15852

  Query Match             4.5%;  Score 57.8;  DB 3;  Length 205044;
  Best Local Similarity   45.8%;  Pred. No. 0.021;
  Matches  237;  Conservative    0;  Mismatches  277;  Indels    3;  Gaps    1;

Qy        609 ATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCA 668
              |||     | ||   |||||| ||    || |  |||||   |    || | |    |  | |
Db     201081 ATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTAT 201022

Qy        669 TTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGT 728
              ||| ||    |   |  || || ||       | ||| | ||||  |||      |  |
Db     201021 ATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTT 200962

Qy        729 GGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCA 788
              || | |    | ||||    |||| |  |     ||  |        ||||   |
Db     200961 ATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTT 200902

Qy        789 ATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCA 848
              |||| | | |    | |||||  |     | || | || | ||||||||  |    | | |  |
Db     200901 TTATATATTTATATTTAAAATAT---ATATATTTTTATATATTTATATTTAAAATATATA 200845

Qy        849 AGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAA 908
              ||     | |||  |   |||  | |     |  ||  | |      |    |    |
Db     200844 TATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATA 200785

Qy        909 AATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTAT 968
              || |      |  |||| ||  |     | |||||  |    |  ||  |  ||||||||
Db     200784 TATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTAT 200725

Qy        969 TTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTC 1028
              | |  | |  |   | | |||||||   || |        | ||  ||    |  | |
Db     200724 ATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATT 200665

Qy       1029 TATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTA 1088
              | | ||| | |  |   |||   ||| || | | |  |    ||   | | | |  ||
Db     200664 TTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATA 200605

Qy       1089 TGCAATGTTATTGCAAAGTATGAATTATTTAAAAATA 1125
              |  ||   ||||   |  |||  ||  || ||| |||
Db     200604 TATAAATATATTTTTATATATTTATATTTAAAATATA 200568


RESULT 13
US-09-949-016-15853/c
; Sequence 15853, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15853
;   LENGTH: 205044
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(205044)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15853

  Query Match             4.5%;  Score 57.8;  DB 3;  Length 205044;
  Best Local Similarity   45.8%;  Pred. No. 0.021;
  Matches  237;  Conservative    0;  Mismatches  277;  Indels    3;  Gaps    1;

Qy        609 ATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCA 668
              |||     | ||   |||||| ||    || |  |||||   |    || | |    |  | |
Db     201081 ATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTAT 201022

Qy        669 TTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGT 728
              ||| ||    |   |  || || ||       | ||| | ||||  |||      |  |
Db     201021 ATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTT 200962

Qy        729 GGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCA 788
              || | |    | ||||    |||| |  |     ||  |        ||||   |
Db     200961 ATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTT 200902

Qy        789 ATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCA 848
              |||| | | |    | |||||  |     | || | || | ||||||||  |    | | |  |
Db     200901 TTATATATTTATATTTAAAATAT---ATATATTTTTATATATTTATATTTAAAATATATA 200845

Qy        849 AGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAA 908
              ||     | |||  |   |||  | |     |  ||  | |      |    |    |
Db     200844 TATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATA 200785

Qy        909 AATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTAT 968
              || |      |  |||| ||  |     | |||||  |    |  ||  |  ||||||||
Db     200784 TATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTAT 200725

Qy        969 TTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTC 1028
              | |  | |  |   | | |||||||   || |        | ||  ||    |  | |
Db     200724 ATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATT 200665

Qy       1029 TATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTA 1088
              | | ||| | |  |   |||   ||| || | | |  |    ||   | | | |  ||
Db     200664 TTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATA 200605

Qy       1089 TGCAATGTTATTGCAAAGTATGAATTATTTAAAAATA 1125
              |  ||   ||||   |  |||  ||  || ||| |||
Db     200604 TATAAATATATTTTTATATATTTATATTTAAAATATA 200568


RESULT 14
US-09-949-016-12387/c
; Sequence 12387, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
```

```
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12387
;   LENGTH: 223471
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(223471)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12387

  Query Match             4.5%;  Score 57.8;  DB 3;  Length 223471;
  Best Local Similarity   45.8%;  Pred. No. 0.022;
  Matches  237;  Conservative    0;  Mismatches  277;  Indels    3;  Gaps    1;

Qy        609 ATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCA 668
              |||     | ||   |||||| ||   || |  |||||  |   || | |   | | |
Db     171508 ATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTAT 171449

Qy        669 TTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGT 728
              |||  ||     |   |  ||  || ||       | |||  | ||||   |||       | |
Db     171448 ATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTT 171389

Qy        729 GGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCA 788
              || | |    |  ||||    |||| |  |    ||  |       ||||   |
Db     171388 ATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTT 171329

Qy        789 ATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCA 848
              ||||  | | |    | |||||  |     | ||  | ||  | ||||||||  |    | | |  |
Db     171328 TTATATATTTATATTTAAAATAT---ATATATTTTTATATATTTATATTTAAAATATATA 171272

Qy        849 AGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAA 908
                ||     | |||  |    |||  | |       |   ||  | |       |   |
Db     171271 TATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATA 171212

Qy        909 AATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTAT 968
              || |       |  |||| ||  |       | |||||  |   |  || |   ||||||||
Db     171211 TATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTAT 171152

Qy        969 TTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTC 1028
              | |  | |  |   | | |||||||   || |         | ||  ||    |  | |
Db     171151 ATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATT 171092

Qy       1029 TATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTA 1088
              |  | ||| | |  |   |||    ||| || | | |  |    ||   | | | |  ||
Db     171091 TTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATA 171032

Qy       1089 TGCAATGTTATTGCAAAGTATGAATTATTTAAAAATA 1125
              |  ||   ||||   |  |||  ||  || ||| |||
Db     171031 TATAAATATATTTTTATATATTTATATTTAAAATATA 170995


RESULT 15
US-09-949-016-12724/c
; Sequence 12724, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12724
;   LENGTH: 223471
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(223471)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12724

  Query Match             4.5%;  Score 57.8;  DB 3;  Length 223471;
  Best Local Similarity   45.8%;  Pred. No. 0.022;
  Matches  237;  Conservative    0;  Mismatches  277;  Indels    3;  Gaps    1;

Qy        609 ATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCA 668
              |||     | ||   |||||| ||   || |  |||||  |   || | |   | | |
Db     171508 ATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTTAT 171449

Qy        669 TTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGT 728
              |||  ||     |   |  ||  || ||       | |||  | ||||   |||       | |
Db     171448 ATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTTTT 171389

Qy        729 GGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCA 788
              || | |    |  ||||    |||| |  |    ||  |       ||||   |
Db     171388 ATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATATATATTT 171329

Qy        789 ATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCA 848
              ||||  | | |    | |||||  |     | ||  | ||  | ||||||||  |    | | |  |
Db     171328 TTATATATTTATATTTAAAATAT---ATATATTTTTATATATTTATATTTAAAATATATA 171272

Qy        849 AGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAA 908
                ||     | |||  |    |||  | |       |   ||  | |       |   |
Db     171271 TATTTTTATATATTTATATTTAAAATATATATATTTTTATATATTTATATTTAAAATATA 171212

Qy        909 AATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTAT 968
              || |       |  |||| ||  |       | |||||  |   |  || |   ||||||||
Db     171211 TATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTAT 171152

Qy        969 TTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTC 1028
              | |  | |  |   | | |||||||   || |         | ||  ||    |  | |
Db     171151 ATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATATATAAATATATT 171092

Qy       1029 TATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTA 1088
              |  | ||| | |  |   |||    ||| || | | |  |    ||   | | | |  ||
Db     171091 TTATATATTTATATTTAAAATATATATAAATATATTTTATATATTTATATTTAAAATATA 171032

Qy       1089 TGCAATGTTATTGCAAAGTATGAATTATTTAAAAATA 1125
              |  ||   ||||   |  |||  ||  || ||| |||
Db     171031 TATAAATATATTTTTATATATTTATATTTAAAATATA 170995


Search completed: May 12, 2006, 03:51:57
Job time : 181.843 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:31:18 ; Search time 827.889 Seconds
(without alignments)
12915.142 Million cell updates/sec

Title: US-09-423-093B-4_COPY_12762_14054
Perfect score: 1293
Sequence: 1 gtgaaagttcaattgttaaa.....ttaaggcaaataagggataa 1293

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1293	100.0	22080	3	US-09-423-093A-4	Sequence 4, Appli
2	82.4	6.4	3673778	6	US-10-312-841-1	Sequence 1, Appli
3	81.2	6.3	7597	6	US-10-311-455-986	Sequence 986, App
c 4	74.8	5.8	158001	7	US-10-211-179-11	GENERAL INFORMATI
5	71.8	5.6	13511	6	US-10-311-455-254	Sequence 254, App
6	71	5.5	15548	6	US-10-311-455-2128	Sequence 2128, Ap
7	71	5.5	61020	7	US-10-221-714A-514	Sequence 514, App
c 8	69	5.3	220	9	US-10-496-207-59	Sequence 59, Appl
9	69	5.3	5449	6	US-10-311-455-1614	Sequence 1614, Ap
10	68.2	5.3	115218	8	US-10-278-698-255	Sequence 255, App
11	68.2	5.3	115218	8	US-10-278-698-769	Sequence 769, App
12	68	5.3	3683	8	US-10-473-126-339	Sequence 339, App
13	67.8	5.2	6292	7	US-10-221-714A-461	Sequence 461, App
14	67.6	5.2	7728	6	US-10-311-455-50	Sequence 50, Appl
15	65.6	5.1	3673778	6	US-10-312-841-2	Sequence 2, Appli
16	65	5.0	7990	6	US-10-311-455-131	Sequence 131, App
17	65	5.0	7990	6	US-10-240-452-7	Sequence 7, Appli
18	64.6	5.0	5739	6	US-10-311-455-691	Sequence 691, App
19	64.6	5.0	83391	7	US-10-433-793-124	Sequence 124, App
20	64	4.9	6175	6	US-10-311-455-1280	Sequence 1280, Ap
21	63.6	4.9	7025	7	US-10-257-166-141	Sequence 141, App
22	63.6	4.9	7025	7	US-10-240-454-45	Sequence 45, Appl
23	63.4	4.9	198522	5	US-10-087-192-244	Sequence 244, App
24	62.8	4.9	7143	6	US-10-311-455-956	Sequence 956, App
25	62.8	4.9	16033	6	US-10-311-455-1377	Sequence 1377, Ap
26	62.4	4.8	5689	5	US-10-239-676-90	Sequence 90, Appl
27	62.4	4.8	5689	6	US-10-240-453-100	Sequence 100, App
28	62.4	4.8	5689	7	US-10-221-714A-148	Sequence 148, App
29	62.4	4.8	6775	7	US-10-433-793-189	Sequence 189, App
30	62.2	4.8	5276	6	US-10-311-455-123	Sequence 123, App
31	62.2	4.8	7346	6	US-10-311-455-318	Sequence 318, App
32	62	4.8	5867	7	US-10-433-793-182	Sequence 182, App
33	62	4.8	12590	6	US-10-311-455-1893	Sequence 1893, Ap
34	61.8	4.8	2437	6	US-10-311-455-2323	Sequence 2323, Ap
35	61.8	4.8	6032	7	US-10-240-454-6	Sequence 6, Appli
36	61.8	4.8	8759	8	US-10-473-126-250	Sequence 250, App
37	61.8	4.8	8759	8	US-10-473-126-378	Sequence 378, App
38	61.6	4.8	6270	6	US-10-311-455-1846	Sequence 1846, Ap
39	61.6	4.8	18154	6	US-10-311-455-227	Sequence 227, App
40	61.4	4.7	5379	6	US-10-311-455-1649	Sequence 1649, Ap
41	61.4	4.7	5379	6	US-10-240-485-129	Sequence 129, App
42	60.8	4.7	6363	6	US-10-240-453-231	Sequence 231, App
43	60.8	4.7	9964	6	US-10-311-455-71	Sequence 71, Appl
44	60.8	4.7	11805	6	US-10-311-455-1721	Sequence 1721, Ap
45	60.8	4.7	19634	8	US-10-473-126-302	Sequence 302, App

ALIGNMENTS

RESULT 1
US-09-423-093A-4
; Sequence 4, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 22080
; TYPE: DNA
; ORGANISM: Salmonella enterica
US-09-423-093A-4

Query Match 100.0%; Score 1293; DB 3; Length 22080;
Best Local Similarity 100.0%; Pred. No. 1.4e-243;
Matches 1293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GTGAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12762 GTGAAAGTTCAATTGTTAAAAATTCCGAGTCATTTAATTGTTGCAGGTTCATCATGGTTA 12821

Qy         61 TCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12822 TCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTTCTATGCTA 12881

Qy        121 GGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12882 GGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCT 12941

Qy        181 GTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12942 GTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGAGCCAAAAAC 13001

Qy        241 AAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13002 AAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTT 13061
```

```
Qy        301 TTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13062 TTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTT 13121

Qy        361 CATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13122 CATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGT 13181

Qy        421 TCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13182 TCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAA 13241

Qy        481 GCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13242 GCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATAC 13301

Qy        541 TATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13302 TATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGT 13361

Qy        601 ATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13362 ATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACA 13421

Qy        661 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13422 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 13481

Qy        721 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13482 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 13541

Qy        781 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13542 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 13601

Qy        841 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13602 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 13661

Qy        901 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13662 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 13721

Qy        961 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13722 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 13781

Qy       1021 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13782 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 13841

Qy       1081 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13842 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 13901

Qy       1141 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13902 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 13961

Qy       1201 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13962 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTTTTTTGGGGGCTTCCA 14021

Qy       1261 CTAACTTACTTAATTAAGGCAAATAAGGGATAA 1293
              ||||||||||||||||||||||||||||||||
Db      14022 CTAACTTACTTAATTAAGGCAAATAAGGGATAA 14054
```

RESULT 2
US-10-312-841-1
; Sequence 1, Application US/10312841

```
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3294164)
US-10-312-841-1

  Query Match             6.4%;  Score 82.4;  DB 6;  Length 3673778;
  Best Local Similarity   45.1%;  Pred. No. 8e-05;
  Matches  390;  Conservative    0;  Mismatches  466;  Indels    8;  Gaps    2;

Qy        272 TACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGG 331
              || || | |  | ||||  |  ||| | ||||||  | ||||||||| | |  |
Db    1713470 TATATTTTATATATATATTTTATATATATTTTATATATATTTTATATATATAATTTAT 1713529

Qy        332 TAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGC 391
              || |    || |||| |||  | | |  ||   |||        | |    |
Db    1713530 ATATATATTTTATATATATTTTATATATATATTTTATATATATTTTATATATTTTATATA 1713589

Qy        392 TCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAA 451
              | |||| ||  |  | | |  ||| |  | ||||  |             |  | | | |
Db    1713590 TATTTTATATATTTTATATATATTTTATATATATTTTATATATTTTATATATATTTTATA 1713649

Qy        452 TACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGA 511
              ||  ||||  | | |  |  | ||  | |  |    ||   |||  |
Db    1713650 TATATTTTATATATATTTTATATATTTTATATATATTTTATATATTTTATATTATA 1713709

Qy        512 TAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTAT 571
              ||  |||  |    |  |||||||||| ||||     ||   |  |    || |  |||
Db    1713710 TATATATAATATATATATTATATATATATTATATATATATTATATATATTTTATATATAT 1713769

Qy        572 CACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGAT 631
                |||  |         |    |  | | || |  | |  |||||| |||||| ||
Db    1713770 TTTATATATTTTATATATATTTTATATATTTTATATATATTTTATATATTTTATATATAT 1713829

Qy        632 ACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGAT 691
              |  | | | |||  | |  | | | | | |  |  |||||||  ||||||  || ||
Db    1713830 ---TTTATTTATATATTTATATATATATATATATTTTATATATATATTTTATATATAT 1713886

Qy        692 CTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGG 751
              |    |||| | | | |  | |||  |  |||    |  |  | | | ||| |||
Db    1713887 TATATATATTTTATATATATTATATATATTTTATATATATTTTATATATATATTTTATAT 1713946

Qy        752 TCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTT 811
              ||||  | |         |  | | |  | ||||  ||  ||  |  | |
Db    1713947 ATATTTTATATATATATTTTATATATATTTTATATATATATTTTATATATATTTTATATA 1714006

Qy        812 TTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTG 871
              |   |||  | | | ||  | ||||     | ||  |  |     |||    |||||  |  |
Db    1714007 TATATTTTATATATATTTTATATATATATTTTATATA-----TATTTTATATATATTTTA 1714061

Qy        872 AATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGC 931
              ||  | | |  |    | | | | |  |  |||      | |  |||| ||
Db    1714062 TATATATATTTTATATATATTTTATATATATATTTTATATATATTTTATATATATATTTT 1714121

Qy        932 TTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATAT 991
              |   |   |      | ||  || || |  |||| ||| ||||| |   | |  |
Db    1714122 ATATATATTTTATATATATTTTATATATATATTTTATATATATATTTTATATATATTTTA 1714181

Qy        992 TTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTG 1051
              | |  | |||  |       || ||   |  | | | || ||| | || | |  |||
```

```
Db    1714182 TATATTTTATATATATTTTATATATTTTATATATATATTTTATATATATTTTATATATTT 1714241

Qy       1052 GCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGA 1111

Db    1714242 TATATATATTTTATATATATTTTATATATATATTTTATATATATATTTTATATATTTAT 1714301

Qy       1112 ATTATTTAAAAATACTTTGGATAT 1135

Db    1714302 ATATTTTATATATATTTTATATAT 1714325


RESULT 3
US-10-311-455-986
; Sequence 986, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 986
;   LENGTH: 7597
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-986

  Query Match             6.3%;  Score 81.2;  DB 6;  Length 7597;
  Best Local Similarity   43.0%;  Pred. No. 1.3e-05;
  Matches  462;  Conservative    0;  Mismatches  603;  Indels    9;  Gaps    1;

Qy         47 GTTCATCATGGTTATCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATC 106

Db        209 GTTTATTATTTATTTATTTATTTATTTATGAATTTATTTATTTATTTATTTATTTATTTA 268

Qy        107 TTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGGTTTATTAG 166

Db        269 TTTATTTATTTATTTATTTATGTATTTATTTATTATTTATTTATTTTATTATTTATTTAT 328

Qy        167 TATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAAT 226

Db        329 TTATTTATTTATTTATTTATGTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 388

Qy        227 GCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTA 286

Db        389 TTAATTATTTATGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTTA 448

Qy        287 TAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAAT 346

Db        449 TTATTTATTTATTTATTTATTTATTTATGTATTTATTTATTTATGTATTTATTTATTTAT 508

Qy        347 ATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCAT 406

Db        509 GTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTTATTATTTAT 568

Qy        407 GTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAAT 466

Db        569 TTATTTATTTATTTATTTATGTATTTATTTATTTATGTATTTATTTATTTATTTACGTAT 628

Qy        467 TGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCT 526

Db        629 TTATTTATTTATTTATTTATTTATTTATGTGTTTATTTATTTATTTATTTATTTATTTGT 688

Qy        527 TGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGT 586

Db        689 TTTATTATTTATTTATTTACGTATTTATATATTTATTTATTTATTTTATTATTCGTTTAT 748

Qy        587 ATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATC 646

Db        749 TTATTTACGTATTTATTTATTTATTTATTTATTTTTTATTTATTTATTTATTTATTTATT 808

Qy        647 ATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTC 706

Db        809 TATTTATTTATTTATTTATTAATTTATGTATTAATTTATGTATTTATTTTATTATTTATT 868

Qy        707 TTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGC 766

Db        869 TATTTATTCGTTTATTCGTTTATTTATTTATTTATTTATGTATTAATTTATTTATTTATT 928

Qy        767 TAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTT 826

Db        929 TATTTATTTATTTTATTATTTATTTATTTATTTATTTATTTATGTATTTATTAGTTTATT 988

Qy        827 TTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAAC 886

Db        989 TATTTATTTATTTATTTATGTATTTATTTATGTATTTATGTATTTATTAATTTATTATTT 1048

Qy        887 AGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATG 946

Db       1049 ATTTATTTAATTATTTATTTATTTATTTATTTAGTTATTAATTTATTTATTTATTTACGT 1108

Qy        947 TTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCA 1006

Db       1109 ATTTATTTATTTATATATTTATTTATTTATTT---------ATTTATTTATTTATCGATA 1159

Qy       1007 AAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTA 1066

Db       1160 TATTTATTTAGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATATATATATA 1219

Qy       1067 TTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAA 1120

Db       1220 TATGTATTTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTGTTTA 1273


RESULT 4
US-10-211-179-11/c
; GENERAL INFORMATION:
;  APPLICANT: Nicholas M. Dean
;  APPLICANT:  Kenneth W. Dobie
;  TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOTYROSYL PHOSPHATASE ACTIVATOR EXPRI
;  FILE REFERENCE: PTS-0011
;  CURRENT APPLICATION NUMBER: US/10/211,179
;  CURRENT FILING DATE:  2002-08-01
;  NUMBER OF SEQ ID NOS: 119
;   OTHER INFORMATION: n = a, t, c, or g
US-10-211-179-11

  Query Match             5.8%;  Score 74.8;  DB 7;  Length 158001;
  Best Local Similarity   44.7%;  Pred. No. 0.00075;
  Matches  388;  Conservative    0;  Mismatches  467;  Indels   13;  Gaps    2;

Qy        255 ATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATT 314

Db      23024 ATATATTATATATTATATATTATATATTTTTATATATTATATATTATATATTATATATTT 22965

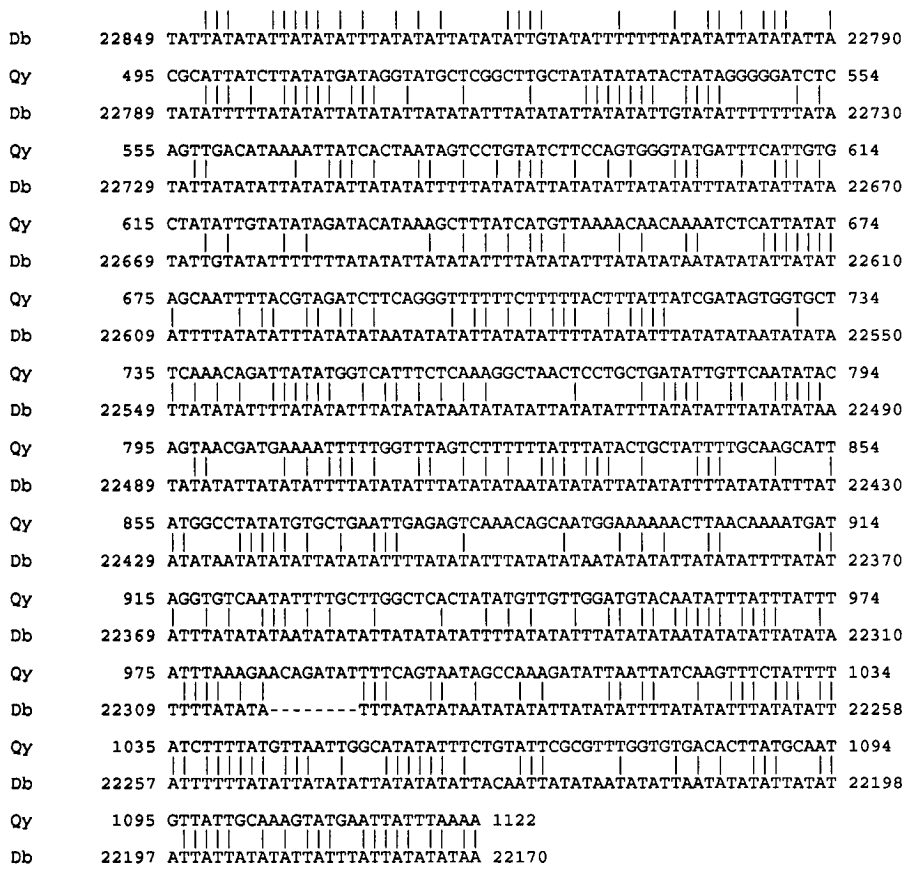
Qy        315 TTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACA 374

Db      22964 TTATATATTAT-----ATATTATATATTATATATTTTTATATATTATATATTATATATTT 22910

Qy        375 GGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGG 434

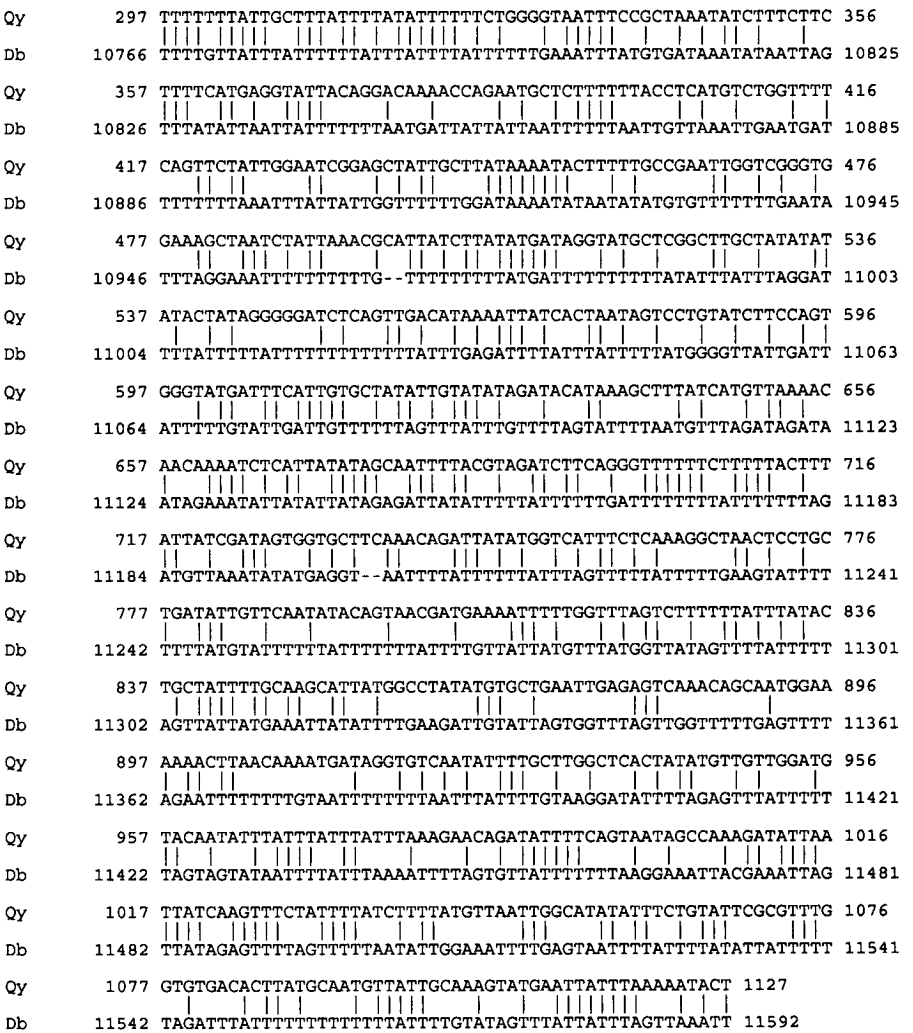
Db      22909 ATATATTATATATTGTATATATTTTTATATATTATATATTATATATTATATATTTTATA 22850

Qy        435 AGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAA 494
```



RESULT 5
US-10-311-455-254
; Sequence 254, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 254
; LENGTH: 13511
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 8889
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-254

Query Match 5.6%; Score 71.8; DB 6; Length 13511;
Best Local Similarity 44.5%; Pred. No. 0.0011;
Matches 370; Conservative 0; Mismatches 457; Indels 4; Gaps 2;



RESULT 6
US-10-311-455-2128
; Sequence 2128, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:

```
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2128
;   LENGTH: 15548
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2128

  Query Match             5.5%;  Score 71;  DB 6;  Length 15548;
  Best Local Similarity   44.9%;  Pred. No. 0.0017;
  Matches  470;  Conservative    0;  Mismatches  570;  Indels    7;  Gaps    5;

Qy         96 TATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTTAAC 155
              ||| |  ||| ||  || |||  |        | |  |||   || ||||  || |
Db      11675 TATATTTTATATTTATTTTATTTTTATATTTTATATTTATTTTATTTTTATATTTTATTT 11734

Qy        156 TGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAATTA 215
              |  | | ||| |||  | |      | ||   |||     |   ||    | |   | |||
Db      11735 TTATATTTTATTATTATATTTTATTTTTATATTTTATTTTTATATTTTATTTTTATATTT 11794

Qy        216 TATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTACA 275
              ||| |    ||       | |    |  ||   ||      | ||| ||      ||  |
Db      11795 TATTTTTATATTTATTTTTATATTTTATTTTTATATTTATTTTTATATTTTATTTTTA 11854

Qy        276 TCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAAT 335
              | | |  |||||  |  ||||| |||||||  | ||||||||||| || |     | ||
Db      11855 TTTTATTTTTATATTTATATTTATTTTATTTTTATATTTTATATTTATTTTATTTTTAT 11914

Qy        336 TTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTT 395
                |    |   ||| |||  | |||  |||   | ||         | |  | | |    ||
Db      11915 ATTTTATTTTTATATTTTATTTTTATATGTTATTTTTATATGTTATA-TATATTTTATTT 11973

Qy        396 TTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACT 455
              || ||  | || | |  |||    || |||   |     |  | |  |||||   || |
Db      11974 TTATATGTTATATATATTTTATTTTTATATGTTATATATATTTTATTTTTATATTTTATT 12033

Qy        456 TTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGG 515
              |||    |   |  |    ||  |   | || | |||     ||     |  | | ||| |
Db      12034 TTTATATGTTATTTTTATATGTTATTTTTATATGTTATTTTTATATGTTATTTTTATATG 12093

Qy        516 TATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACT 575
              |    |      || ||| | ||||   ||      ||   |  | | ||| |  | | | |
Db      12094 TTATTTTTATATGTTATTTTTATATGTTATTTTTAT---ATGTTATATATATATGTTATT 12150

Qy        576 AATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACAT 635
              ||    | ||| |    ||       || |  |     |||||  |||||| |||  |
Db      12151 TTTATAGTTATATATATATGTTATATATATATGTTATATATATATGTTATATATATATGT 12210

Qy        636 AAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTC 695
              |     |      |  || |   | |  | |  |||||   |  | ||  ||  |  |
Db      12211 TATATATATATGTTATATATATATATGTTATATATATATGTTATATATATATATGTTATA 12270

Qy        696 AGGGTTTTTTCTTTTTA-CTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCA 754
                | | || | | ||  | | ||||  |||   ||  |  | | | | ||  |||    |
Db      12271 TATATATGTTATATATATATATGTTATATATATATGTTATATATATATATGTTATATATA 12330

Qy        755 TTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTG 814
              | | | | |     |    | |     | |   ||||| |   | | |  | || | |
```

```
Db      12331 TATGTTATATATATATATGTTATATATATATGTTATATATATATATGTTATATATATATA 12390

Qy        815 GTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAAT 874
                ||| | | | |   |||||    ||| ||  | | ||     |          |  |
Db      12391 TGTTA-TGTATATATGTTATATATATATATTTTATGTATATATGTTATATATATATATTT 12449

Qy        875 TGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTG 934
              |   |  |  |        ||  | | |  |||    | ||  |    | |  |||| ||   ||
Db      12450 TATGTATATATTTTATATATATATATTTTATGTATATATTTTATATATATATATTTTATG 12509

Qy        935 GCT-CACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTT 993
                 |  | | ||| |    |  || |     ||| |||||  |    |  |           |
Db      12510 TATATATTTTATATATATATATTTTATGTATATATTTTATATATATATATTTTATGTATA 12569

Qy        994 TCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGC 1053
              |    | |||   | | ||  |||  |||  |  || ||| |      ||     ||  |  |
Db      12570 TATTTTATATATATATATTTTATGTATATATGTTATATATATATATATATATATTTC 12629

Qy       1054 ATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAAT 1113
                |||||  | | | ||     |    |   | | ||||    | |   |  |   ||| ||
Db      12630 GTATATATATATTTTATATATATATTATATATTTATTTCGTATATATATATTTTATATAT 12689

Qy       1114 TATTTAAAAATACTTTGGATATTAGTA 1140
                 ||| | |||  ||   |||   ||
Db      12690 ATATTATATATATATTTCGTATATATA 12716


RESULT 7
US-10-221-714A-514
; Sequence 514, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 514
;   LENGTH: 61020
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-514

  Query Match             5.5%;  Score 71;  DB 7;  Length 61020;
  Best Local Similarity   44.9%;  Pred. No. 0.0029;
  Matches  484;  Conservative    0;  Mismatches  580;  Indels   15;  Gaps    5;

Qy         34 TTAATTGTTGCAGGTTCATCATGGTTATCCAAAATAATAATTGCCGGGGTGCAGTTAGCA 93
              | | || | |  ||  | |   |||      ||| || ||          ||
Db      33944 TAAGTTATAGAAATTTTTTTAATTTTAGTTTTTATATTAGTTAAATTATATTTTTTTTAT 34003

Qy         94 AGTATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTTA 153
              || |||| || | || | | | |   | |  | ||  |||||  | | |  |  | ||
Db      34004 AGGATTTTATTTTTTTTGTGTTTTTAATGAAATGATGAATATAAATTATGAATATCGTGT 34063
```

```
Qy        154 ACTGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAAT 213
              ||  |   | || ||||  |  |  || | | ||   ||   ||  |||     |  ||
Db      34064 TTTGTATATATTGTTTGGTTAAAAG--GTAGTTATTATTATTATTATTGGATATTATTAT 34121

Qy        214 TATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTA 273
              |||   | | |   |     |    | ||    |    | ||   | |    |||
Db      34122 TATTATTGAAATTAATTAAATTTTGATTTTATTTTGATTAATGGTTTTTTTTTATTTTTA 34181

Qy        274 CATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTA 333
              || |||   | ||   |  |  || |||||||   |||||||  ||| |||| | |  ||
Db      34182 AATGTAA--ATAATTATTTGTTGTTATTTTATTTTGTTATTTTTTATATTTTATTGTTTA 34239

Qy        334 ATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTC 393
              |     | |||| ||| || || |   |   || ||   | | |  |          |
Db      34240 TTCGTTTTTTAATATTTTTGTTTTTGTTTTTTAGTTTTTTTGTAAATATTTTTTGTAAGT 34299

Qy        394 TTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATA 453
              |||  |   | | | |  | | || | ||  | | |    | ||| ||||||  |
Db      34300 TTTAATTTGTTTTATTTTATGGTTATTTTTTTTTAAGTTGTTTTTTTTGTTTATAATGTT 34359

Qy        454 CTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATA 513
              |||       |||  |   |  ||    |      ||       | | || || ||||
Db      34360 TTTTCGATTTTATTATTATTTTTAAATTTAATAAATTT----TTTATTTTTTTTAAGATA 34415

Qy        514 GGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCA 573
              | |  |    || |   | |||| || |    || |   | |  | | ||
Db      34416 TTTTTAAGGATATTTTTTGGTTTATAGTAAAATAAAATATAGAGTTATTTTATATATATT 34475

Qy        574 CTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATAC 633
              |||| | | | |  | |  ||| |||     |||  |||     || || ||
Db      34476 TATTTAGTTATATGTATAGTATTTTTTATTATTAATATGTTTTATTAGAAATGTATATTT 34535

Qy        634 ATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCT 693
              ||  |  |||||  |  ||| | | |   | | | ||||| ||    || ||  |    |
Db      34536 ATTATAATTTATGAATTTATATCGATGTATTATTATTATTTAAAGTTTATAGTTTATATT 34595

Qy        694 TCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTC 753
              || | |||| | | |  | |||| |       | |             || |||||
Db      34596 AGGGGTTATTTTTTGTGTTGTATATTTTATGGGTTTGGATAAATGGATAATGATATGTAT 34655

Qy        754 ATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACA-----GTAACGATGAAAA 808
              | |   |  |   |   |  ||||   | | | ||  |     | | | | |
Db      34656 TTATTATTATAGTATTATATAGAATATATATATTATTTTAAAAATTTTTTATGTTTTATT 34715

Qy        809 TTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTG 868
              | |||  |||  |  ||||  ||| |  || || ||    |   | | | | | | |
Db      34716 TATTTATTTTTTTTGTTTTATTTTTTTTTGATAATTATTGATTTTATTAGTATTTTTATA 34775

Qy        869 CTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTT 928
              |    |   | |       |||||  | | |   | ||  |   |     |||||
Db      34776 GTTTTGTTTTTTTTAGTATATGTTGGAATTATATAATATATATTTTTTTTAGATTTATTT 34835

Qy        929 TGCTTGGCTCACTATATGTTGTT--GGATGTACAATATTTATTTATTTATTTAAAGAACA 986
              |  ||    |  |||||  ||  |  | |   || ||| |||||   || | ||   |
Db      34836 TTTTTATTTAGTAATATGTATTTAAGTTTTTTTTATGTTTTTTTATGGTTTGATAGTTTA 34895

Qy        987 GATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTT 1046
              | |||| |||  |   ||  | ||   | |  |  || || |||||   ||   ||
Db      34896 TTTTTTTTTAGTTTTGAATAATATTTTTTTGTTTAGATATTATAGTTTATTAGTTTATTT 34955

Qy       1047 AATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAA 1105
              | |     |||  | ||  ||  |  ||  |    |  || ||| ||  |||
Db      34956 ATTGAAGGATATTTTAGTTGTTTTCGAGTTTTGGTAATTATGAATAAAGTTGTTATAAA 35014
```

RESULT 8
US-10-496-207-59/c
; Sequence 59, Application US/10496207
; Publication No. US20050079168A1
; GENERAL INFORMATION:
; APPLICANT: Karlyshev, Andrey
; APPLICANT: Wren, Brendan
; TITLE OF INVENTION: Virulence Genes, Proteins, and Their Use
; FILE REFERENCE: GJE-6947
; CURRENT APPLICATION NUMBER: US/10/496,207
; CURRENT FILING DATE: 2004-05-19
; PRIOR APPLICATION NUMBER: PCT/GB02/05212
; PRIOR FILING DATE: 2002-11-18
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 220
; TYPE: DNA
; ORGANISM: Yersinia pseudotuberculosis
US-10-496-207-59

Query Match 5.3%; Score 69; DB 9; Length 220;
Best Local Similarity 58.5%; Pred. No. 0.00084;
Matches 120; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
Qy        644 ATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTT 703
              | | |  || ||  | |     |  |  ||   | ||| | ||| |  | || || |||
Db        209 AGCTTCGTAGAATTAGATCTGATATTATTAAGCCTATTGTCCGTCGTGCCTCTGGTTTTC 150

Qy        704 TTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAA 763
              |  ||||| || | ||| | | | |   |||| || ||||| || || || || |||
Db        149 TAATTTTTTCTATTTTAACTACAATTATTATTCAGACTGATTACATAGTAATGTCCCAAG 90

Qy        764 GGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCT 823
                  |||  |||  ||| ||  |  |||||  ||    ||||| |||||||||  | || |
Db         89 TTTTAAAACCTAGTGAGATAATAACATATAGCGTTTTAATGAAGATTTTTGGTATTGTAT 30

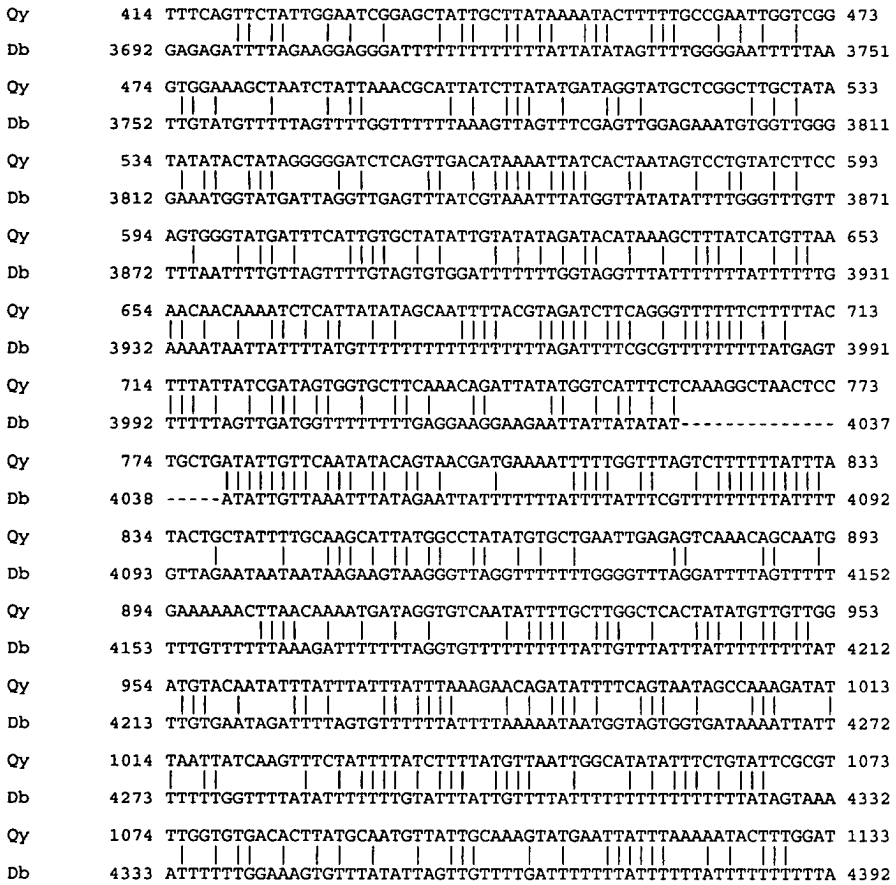
Qy        824 TTTTTATTTATACTGCTATTTTGCA 848
              | ||| | |   ||||  |  | ||
Db         29 TCTTTGTCTTCTCTGCGTTGCTTCA 5
```

RESULT 9
US-10-311-455-1614
; Sequence 1614, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1614
; LENGTH: 5449
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1614

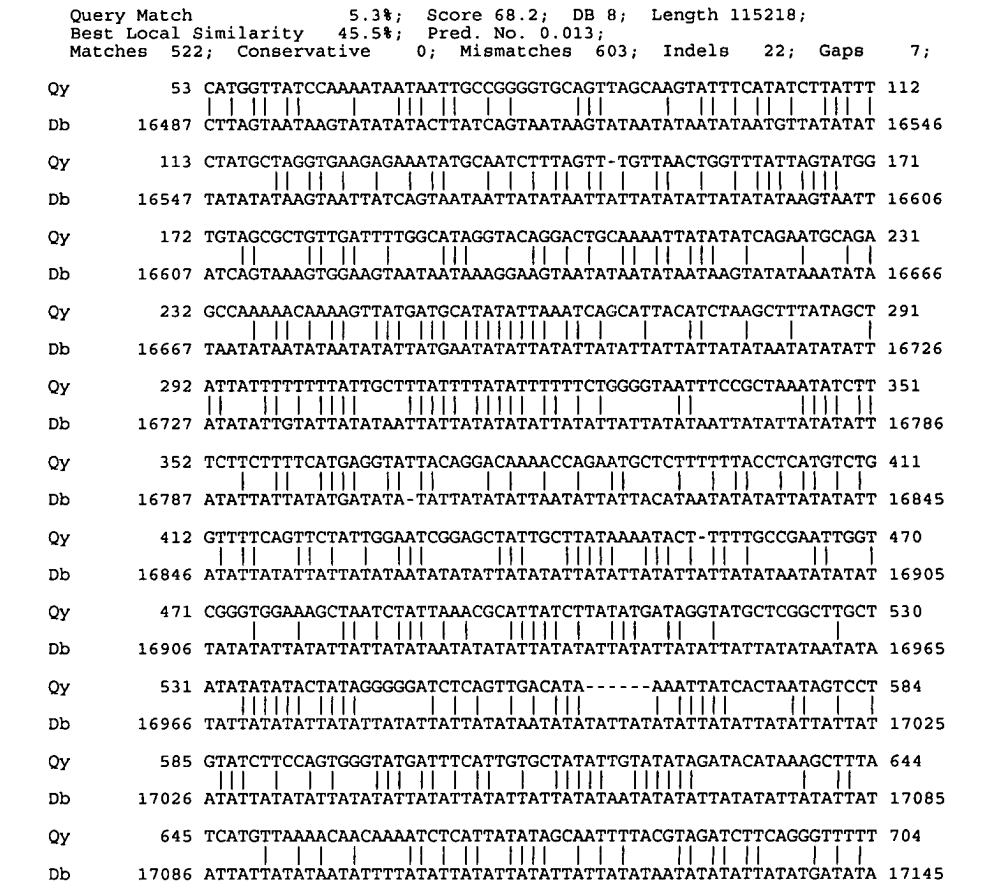
Query Match 5.3%; Score 69; DB 6; Length 5449;
Best Local Similarity 44.7%; Pred. No. 0.0029;
Matches 377; Conservative 0; Mismatches 445; Indels 21; Gaps 2;

```
Qy        296 TTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTT 355
              | |||||  ||  || ||||| ||||||||  |    | ||||    |  || |  | ||
Db       3572 TATTTTTAGTTTTTTGATTTTGTATTTTTTTTTTTTTTATTTGTTTTTGATTTTATGTTT 3631

Qy        356 CTTTTCATGAGGTATTA--CAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGT 413
              ||||  ||   |  |||   || | |  |  |   || | ||||||   |  | | | |
Db       3632 TTTTTTATTTAGGTTTATTTAGAAGATTATTATTTTGGTTTTTTTTTTTATAGTTTTTAGG 3691
```



RESULT 10
US-10-278-698-255
; Sequence 255, Application US/10278698
; Publication No. US20050037344A1
; GENERAL INFORMATION:
; APPLICANT: PathoArray GmbH
; APPLICANT: Stuhlmuller, Bruno
; APPLICANT: Haupl, Thomas
; TITLE OF INVENTION: Nucleic Acid Array
; FILE REFERENCE: O30027US
; CURRENT APPLICATION NUMBER: US/10/278,698
; CURRENT FILING DATE: 2002-10-23
; NUMBER OF SEQ ID NOS: 1050
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 255
; LENGTH: 115218
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-278-698-255



```
Qy       1064 GT-ATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAA 1122
              | ||     | | | | | |  | ||  | |  || |  |  |||| |  ||| | | |
Db      17495 ATAATATATATATTATATAATTATAATATATTAATAATATATCGTATTATATATATTATA 17554

Qy       1123 ATACTTTGGATATTAGTACCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTT 1182
              ||  ||  ||||||     | |    ||   |  |  ||    |  ||  ||| ||
Db      17555 TTATATTATATATTATATTATTAAAATATTATCTATATTATAAATATATTATATTATATT 17614

Qy       1183 TCTAGTA 1189
              | || ||
Db      17615 TTTATTA 17621


RESULT 11
US-10-278-698-769
; Sequence 769, Application US/10278698
; Publication No. US20050037344A1
; GENERAL INFORMATION:
;  APPLICANT: PathoArray GmbH
;  APPLICANT:  Stuhlmuller, Bruno
;  APPLICANT:  Haupl, Thomas
;  TITLE OF INVENTION: Nucleic Acid Array
;  FILE REFERENCE: O30027US
;  CURRENT APPLICATION NUMBER: US/10/278,698
;  CURRENT FILING DATE:  2002-10-23
;  NUMBER OF SEQ ID NOS: 1050
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 769
;   LENGTH: 115218
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-278-698-769

  Query Match             5.3%;  Score 68.2;  DB 8;  Length 115218;
  Best Local Similarity   45.5%;  Pred. No. 0.013;
  Matches  522;  Conservative    0;  Mismatches  603;  Indels   22;  Gaps    7;

Qy         53 CATGGTTATCCAAAATAATAATTGCCGGGGTGCAGTTAGCAAGTATTTCATATCTTATTT 112
              | | ||  ||     |  ||| ||  | |      |||     |  || | ||  |  ||| |
Db      16487 CTTAGTAATAAGTATATATACTTATCAGTAATAAGTATAATATAATATAATGTTATATAT 16546

Qy        113 CTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTT-TGTTAACTGGTTTATTAGTATGG 171
              || || |  |  | ||   | | | || || |  ||  |  | ||| ||||
Db      16547 TATATATAAGTAATTATCAGTAATAATTATATAATTATTATATATTATATATAAGTAATT 16606

Qy        172 TGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAATTATATATCAGAATGCAGA 231
              ||    || ||  |    |||      || | |  || || |||  |     |   | |
Db      16607 ATCAGTAAAGTGGAAGTAATAATAAAGGAAGTAATATAATATAATAAGTATATAAATATA 16666

Qy        232 GCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCT 291
              | ||  | ||  ||| ||| |||||||| || |    |   ||   |  |       |
Db      16667 TAATATAATATAATATATTATGAATATATTATATTATATTATTATTATATAATATATATT 16726

Qy        292 ATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTT 351
              ||   || | ||||    |||||  |||||  || | |      ||          |||| ||
Db      16727 ATATATTGTATTATATAATTATTATATATATTTATATTATTATATAATTATATTATATATT 16786

Qy        352 TCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTG 411
              |  ||  ||||  || ||    | |  |  |  ||    |  | ||  | || | | |
Db      16787 ATATTATTATATGATATA-TATTATATATTAATATTATTACATAATATATATTATATATT 16845

Qy        412 GTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACT-TTTTGCCGAATTGGT 470
              | ||   || |  |  |||      |||   |||||  ||| | || |      ||   |
Db      16846 ATATTATATTATTATATAATATATTATATATTATATTATATTATTATATAATATATAT 16905

Qy        471 CGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCT 530
              |   |   || | ||| | |   ||||| |  |||  ||  |        |
Db      16906 TATATATTATATTATTATATAATATATTATATATTATATTATATTATTATATAATATA 16965

Qy        531 ATATATATACTATAGGGGGATCTCAGTTGACATA------AAATTATCACTAATAGTCCT 584
              |||||| ||||      | | |  | | |||      | |||||     ||  |  |
Db      16966 TATTATATATTATATTATATTATTATATAATATATTATATATTATATTATATTATTAT 17025

Qy        585 GTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTA 644
              |||  |  | |  ||| || | ||  |  ||||  ||||||          |  ||
Db      17026 ATATTATATATTATATATTATATTATTATTATATAATATATTATATATTATATTAT 17085

Qy        645 TCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTT 704
              |  | | |    || | ||  |||| | | |    || || ||    | | |
Db      17086 ATTATTATATAATATTTTATATTATATTATATTATTATATAATATATTATATGATATA 17145

Qy        705 TCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAG 764
              |  | ||| | ||| ||  ||| |      |  |  | |||| ||  |   || |   |
Db      17146 TTATATTATTATATAATATATTATATGATATATTATATTACATTATATATTATATTA- 17204

Qy        765 GCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTT 824
                      ||||||  |  | ||| | ||      |  || | |  | |  | |
Db      17205 ----------TATGATATATTATATGATATATTATATTATATGATATATTATATTATATA 17254

Qy        825 TTTTAT-TTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCA 883
              || |||  |||| |   || |  ||   | |||  | ||||       |  | | |
Db      17255 TTATATAATATAATATAATATATCATATAATATATCATATAATATAATATATCATATAAT 17314

Qy        884 AACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTAT 943
              |      ||  ||||  |  |  | |  |||      ||||| ||  |    |  ||
Db      17315 ATATTATATAATAAAATATATAATATAATATATAATATAATATATTATTATATATATAAT 17374

Qy        944 ATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAG 1003
              || || |  |  ||  |||| |||     ||| |||          || |   || ||
Db      17375 ATATTATATAACATATTATATATATAATAATATATAATATTATATAATATATTATATTAT 17434

Qy       1004 CCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCT 1063
              | |      || ||  |  || |||| |||  | |||  | | |    |||| || |
Db      17435 ATATAATATAATATAATATATTATATTATATTATATATTATTAATATATTATAATTAT 17494

Qy       1064 GT-ATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAA 1122
              | ||     | |  | | |  | ||  | |  || |  |  |||| |  ||| | | |
Db      17495 ATAATATATATATTATATAATTATAATATATTAATAATATATCGTATTATATATATTATA 17554

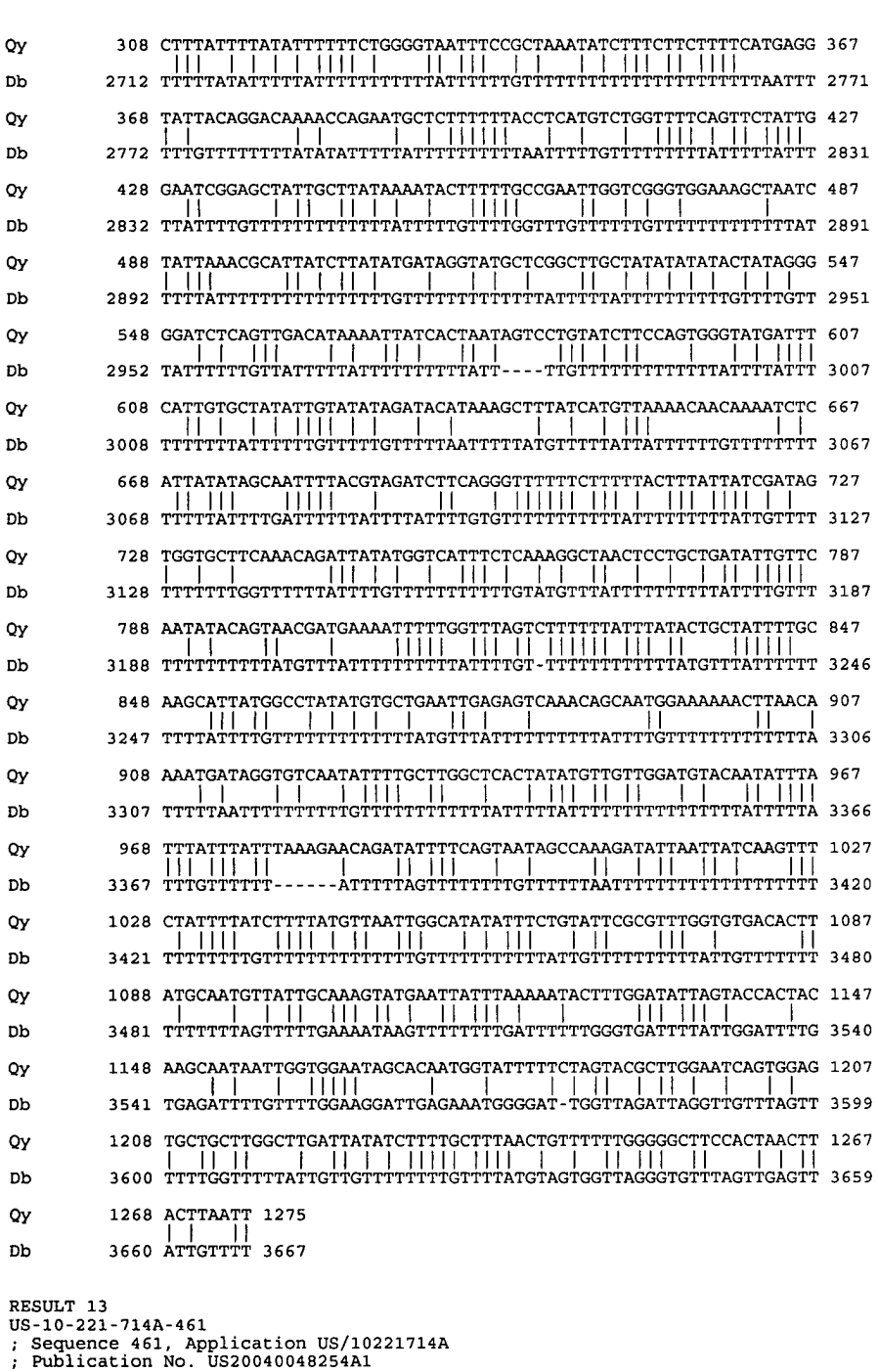
Qy       1123 ATACTTTGGATATTAGTACCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTT 1182
              ||  ||  ||||||     | |    ||   |  |  ||    |  ||  ||| ||
Db      17555 TTATATTATATATTATATTATTAAAATATTATCTATATTATAAATATATTATATTATATT 17614

Qy       1183 TCTAGTA 1189
              | || ||
Db      17615 TTTATTA 17621


RESULT 12
US-10-473-126-339
; Sequence 339, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Methods and nucleic acids for the analysis of  hematopoietic cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/473,126
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 339
;   LENGTH: 3683
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-339

  Query Match             5.3%;  Score 68;  DB 8;  Length 3683;
  Best Local Similarity   44.4%;  Pred. No. 0.0039;
  Matches  456;  Conservative    0;  Mismatches  560;  Indels   12;  Gaps    4;

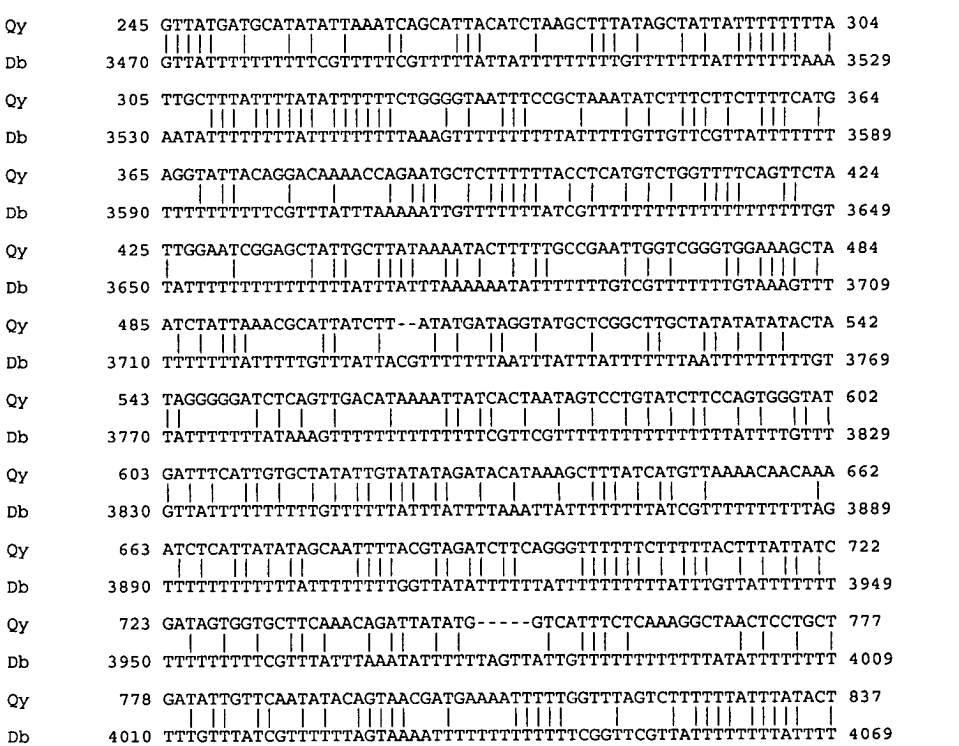
Qy        248 ATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTATTG 307
              | | || |  |||||  |    |||  | | |  |||    | |  |||||  | ||
Db       2652 AGGTTGTAGTTATTATGTATTTATTTATTTTTATTTTTATTTTTTATTTTTTGATTTTT 2711
```



RESULT 13
US-10-221-714A-461
; Sequence 461, Application US/10221714A
; Publication No. US20040048254A1

; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 461
; LENGTH: 6292
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-461

Query Match 5.2%; Score 67.8; DB 7; Length 6292;
Best Local Similarity 43.6%; Pred. No. 0.0052;
Matches 405; Conservative 0; Mismatches 517; Indels 7; Gaps 2;



```
Qy        838 GCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAA 897
Db       4070 TTTTTTATTGGTTATTTTTTTTTTTTTTTTTATTTATTTAAAAGTTTTTTTTTTTATTGT 4129

Qy        898 AAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGT 957
Db       4130 TTTTTTATAAAATTTTTTTTTTTTTTTGTTCGTTATTTTTTTTTTTTTTTTTTTTTGT 4189

Qy        958 ACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAAT 1017
Db       4190 ATTTTTTTTTTTTTTTTTATTTGTTATTTTTTTTTTTTTTTATTATTTAAAATTTTTTT 4249

Qy       1018 TATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGG 1077
Db       4250 ATTTATAGTTTTTTTTTTTTTTTTTTTTTTTTTTTATTATATTTTTGTAAATTTTTTTTTT 4309

Qy       1078 TGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTA 1137
Db       4310 TTTTGTTTATTTTCGTTTTTTTTTTACGATTTTTTTTTATTTTTTTTTATTTATTTAAAA 4369

Qy       1138 GTACCACTACAAGCAATAATTGGTGGAAT 1166
Db       4370 ATTTTTTTTTTATTATTTTTTTGTGAAAT 4398


RESULT 14
US-10-311-455-50
; Sequence 50, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 50
;   LENGTH: 7728
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-50

  Query Match             5.2%;  Score 67.6;  DB 6;  Length 7728;
  Best Local Similarity   42.8%;  Pred. No. 0.0062;
  Matches  457;  Conservative    0;  Mismatches  604;  Indels    7;  Gaps    2;

Qy         86 AGTTAGCAAGTATTTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTA 145
Db       5944 AGGTTTAACGTATTAATTTTTATAGGTATCGGATATTGTAAGTTAAATAAAATTTTTTTA 6003

Qy        146 GTTTGTTAACTGGTTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGAC 205
Db       6004 TTTTTTTATTTTTTATTTGGTAATTGTTAAAGAGAGTTAGTTTTAAAAATTTATATATTT 6063

Qy        206 TGCAAAATTATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAAT 265
Db       6064 TATAATGTTTTGGAAAATGATAGAAAGTTTTAATAATGAGGGTTGATTAGTATATTTGAG 6123

Qy        266 CAGCATTACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTT 325
Db       6124 AGGTCGTGTATTTTTTTTATAGAGTTTTTATTTTAATTTTTTTTTTTTTTTATTTTTTAT 6183

Qy        326 CTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCA 385
Db       6184 TTTTATTTTATATCGGGATTTGTTATGGGTTTTATTTGATTTTTTTGTTTTTGGTTTTTT 6243

Qy        386 GAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTT 445
Db       6244 TTGTTTTATTTAATTTGGTATATTATAATTAAAGGAGTATTTTTAAAATGTTGTTTTGAT 6303

Qy        446 ATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTT 505
Db       6304 TAAGATGTTTTTTTGTTTAGTAATTTTATTTTTTTAATTTAATATTTATAG--TTTTTTT 6361

Qy        506 ATATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAA 565
Db       6362 CGGTTTAATTTTAAATGGTTTTTTTAGTTTTATTTTATATTAGAATTTTTTTTATTATTA 6421

Qy        566 AATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTAT 625
Db       6422 TAATTTGTATTTTTGTATTTTTTTTTTTTTTTGGGGATTTATATTTTTTTTTTTTGTATAT 6481

Qy        626 ATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTAC 685
Db       6482 TTATATTTTTTTATTGGGTTTTTTTTATTAAAATTTTTTTTTGTTTTTTAGCGAATTTAG 6541

Qy        686 GTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATT 745
Db       6542 TTTTGGAGATGTGGTTGGATTTTTTATTTTTTTTGTTTATTGTTTTTGAAAATGATTTTT 6601

Qy        746 ATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGA 805
Db       6602 TTTTTTTTTGATTTATAGTAGTGATTTTGTTAATTTTATTTTAGGTGGTAATTTATAATA 6661

Qy        806 AAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATAT 865
Db       6662 GGTTGTTTTTTGATATTTTTTATGTTGTTATTTAATTTTTATATTTTTTTTTTTTTTTTT 6721

Qy        866 GTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATA 925
Db       6722 TTTATAGATTGTAAGTTTATTGAGAGTTAGAATTGTTTTATATATTTTTA-----TATTT 6776

Qy        926 TTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAAC 985
Db       6777 TTTATGTTTATTGATATTTATTTGATATTAATGAATATTTAGTTTTGATTTATAAATATA 6836

Qy        986 AGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGT 1045
Db       6837 TGATATTTAATTTATTAAAATATTTTTATTTTTATATATTTGAGATATAATTGAGTGTTT 6896

Qy       1046 TAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAA 1105
Db       6897 TAGTAGTTATAATTTTTATTTTTGTTTATAATTTATAGTAATAATTTATATTTTTATTAT 6956

Qy       1106 GTATGAATTATTTAAAAATACTTTGGATATTAGTACCACTACAAGCAA 1153
Db       6957 TTATTTAATATTTTTATATAAATATTTTAAAATATTTTATAAAAGTAA 7004


RESULT 15
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
```

```
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (379615)
US-10-312-841-2

  Query Match             5.1%;  Score 65.6;  DB 6;  Length 3673778;
  Best Local Similarity   45.6%;  Pred. No. 0.16;
  Matches  267;  Conservative    0;  Mismatches  316;  Indels    2;  Gaps    1;

Qy        661 AAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTA 720
              || | |  ||| || | | ||| |  | | | || ||   ||||| | | || ||| | |
Db     351018 AATTTTTTTTAAATCGTATTTTAATTTTGTTTTTGAGTAGTTTTTGTGTATATTTTGTGA 351077

Qy        721 TCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGAT 780
              |  ||||||||| || ||  | ||     |    | | |                |  |  |
Db     351078 TATTTAGTGGTGATTAAATTATATATGTAGTAATTATTTTTTTTTTGTTTGTTTTATTTT 351137

Qy        781 ATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCT 840
                ||  |  || | ||| ||| ||  ||||   |||||  |||| | |||||  |
Db     351138 TGAATTTTAAGTAGATTAAGGATAAAGTTTTTATTTTTAGAGTTTTATTTTTATTATTTA 351197

Qy        841 ATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAA 900
                 ||   ||||      ||| |  | | |  |     |    |    || |
Db     351198 GAGAGTGAAAGTTTATAAAAAGTATTTATTAATTGAATGAAAATTAATATTTGAAGTGTG 351257

Qy        901 CTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACA 960
                  | | | |||| ||| ||| | ||       |      | |   |  || |
Db     351258 GAATTTGTTTTAAAATGTGTTAATTTTTAGTTTTTAGATTTTGTAAATATAATAAATATA 351317

Qy        961 ATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTAT 1020
              |  |   || | ||   | ||  | |  | | |    || |    | |     || || |
Db     351318 AAGTAATTTAAGTTGAATTAATTATATTTGTGTAGTTTAGTTTATATATCGTGTATTTTT 351377

Qy       1021 CAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGT 1080
               ||| | |   || | || |  ||  ||    || |||  |  ||    ||| |||
Db     351378 TTTTTTTTTTTGGGATTTATTTTTGTAGGTGAGTAATTTTTTAGATTTTAAATTTAGTGG 351437

Qy       1081 GACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTA 1140
              |  | || |  | | ||  ||    ||||  | | | |    ||| | || || | ||||
Db     351438 GGGATTTTT--ATTATTTATGTGCGGTATTGAATTTGTTGTTATATTATGTATTTAAGTA 351495

Qy       1141 CCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTGGAATC 1200
                    |  |  || |  | |       |  |||||| |      | ||   |
Db     351496 GAGGATTGGTTTTTTTTATGTTGGGTTGGGAGGGGAAATTTTTATTTAGGGATTATTAAT 351555

Qy       1201 AGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTTTAACTGTT 1245
                || |||| |  |||| || ||| | | ||||   |  | | ||
Db     351556 TATGTAGTGTTATTTGGTTTTATTTTTTTTTTTAGGTCGAATTTT 351600


Search completed: May 12, 2006, 05:35:11
Job time : 840.889 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:52:36 ; Search time 658.783 Seconds
(without alignments)
8004.956 Million cell updates/sec

Title: US-09-423-093B-4_COPY_12762_14054
Perfect score: 1293
Sequence: 1 gtgaaagttcaattgttaaa.....ttaaggcaaataagggataa 1293

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/2/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/2/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	62.4	4.8	171486	17	US-11-121-086-105	Sequence 105, App
	2	60.4	4.7	5152	10	US-10-240-708-73	Sequence 73, Appl
	3	59.6	4.6	173602	17	US-11-121-086-25	Sequence 25, Appl
	4	57.2	4.4	49979	10	US-10-995-561-13443	Sequence 13443, A
c	5	56.4	4.4	990	12	US-10-301-480-568791	Sequence 568791,
c	6	56.4	4.4	990	12	US-10-301-480-1182200	Sequence 1182200,
	7	55.2	4.3	6583	10	US-10-240-708-26	Sequence 26, Appl
c	8	54.2	4.2	171486	17	US-11-121-086-105	Sequence 105, App
	9	54	4.2	6499	17	US-11-011-332A-77	Sequence 77, Appl
c	10	53.6	4.1	607	12	US-10-301-480-602391	Sequence 602391,
c	11	53.6	4.1	607	12	US-10-301-480-1215800	Sequence 1215800,
	12	53.6	4.1	6499	17	US-11-011-332A-91	Sequence 91, Appl
c	13	53	4.1	990	12	US-10-301-480-568793	Sequence 568793,
c	14	53	4.1	990	12	US-10-301-480-1182202	Sequence 1182202,
	15	52.8	4.1	481	12	US-10-301-480-563203	Sequence 563203,
	16	52.8	4.1	481	12	US-10-301-480-1176612	Sequence 1176612,
c	17	52.8	4.1	173602	17	US-11-121-086-25	Sequence 25, Appl
c	18	52.4	4.1	119036	10	US-10-995-561-13314	Sequence 13314, A
	19	52.2	4.0	6669	10	US-10-240-708-6	Sequence 6, Appli
	20	52.2	4.0	10467	10	US-10-240-708-2	Sequence 2, Appli
c	21	52	4.0	920	12	US-10-301-480-558120	Sequence 558120,
c	22	52	4.0	920	12	US-10-301-480-1171529	Sequence 1171529,
c	23	52	4.0	924	12	US-10-301-480-558121	Sequence 558121,
c	24	52	4.0	924	12	US-10-301-480-1171530	Sequence 1171530,
	25	51	3.9	1203	18	US-11-079-463-1993	Sequence 1993, Ap
	26	50.8	3.9	587	12	US-10-301-480-572793	Sequence 572793,
	27	50.8	3.9	587	12	US-10-301-480-1186202	Sequence 1186202,
	28	50.8	3.9	648	12	US-10-301-480-572792	Sequence 572792,
	29	50.8	3.9	648	12	US-10-301-480-1186201	Sequence 1186201,
	30	50.8	3.9	6040	10	US-10-240-708-69	Sequence 69, Appl
c	31	50.4	3.9	201990	10	US-10-995-561-13303	Sequence 13303, A
c	32	50.2	3.9	990	12	US-10-301-480-568794	Sequence 568794,
c	33	50.2	3.9	990	12	US-10-301-480-1182203	Sequence 1182203,
c	34	50.2	3.9	991	12	US-10-301-480-568792	Sequence 568792,
c	35	50.2	3.9	991	12	US-10-301-480-568795	Sequence 568795,
c	36	50.2	3.9	991	12	US-10-301-480-1182201	Sequence 1182201,
c	37	50.2	3.9	991	12	US-10-301-480-1182204	Sequence 1182204,
	38	50	3.9	5152	10	US-10-240-708-74	Sequence 74, Appl
	39	49.8	3.9	28536	17	US-11-011-332A-155	Sequence 155, App
	40	49.4	3.8	20317	10	US-10-995-561-13460	Sequence 13460, A
c	41	49.2	3.8	592	7	US-09-925-065A-689095	Sequence 689095,
	42	49.2	3.8	53332	10	US-10-786-065-3	Sequence 3, Appli
c	43	49.2	3.8	184868	17	US-11-121-086-88	Sequence 88, Appl
	44	49	3.8	26772	10	US-10-995-561-13313	Sequence 13313, A
	45	49	3.8	54946	10	US-10-995-561-13479	Sequence 13479, A

ALIGNMENTS

RESULT 1
US-11-121-086-105
; Sequence 105, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 105
; LENGTH: 171486
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-105

Query Match 4.8%; Score 62.4; DB 17; Length 171486;
Best Local Similarity 45.3%; Pred. No. 0.96;
Matches 475; Conservative 0; Mismatches 551; Indels 22; Gaps 6;

```
Qy        99 TTCATATCTTATTTCTATGCTAGGTGAAGAGAAATATGCAATCTTTAGTTTGTTAACTGG 158
             ||  ||| | || | |     |    | | | |||| | | | |||| |  ||  |
Db     16604 TTATTATATAATATATTAATAAATATATTATATATATTTATTATATAGTATAATATATTA 16663

Qy       159 TTTATTAGTATGGTGTAGCGCTGTTGATTTTGGCATAGGTACAGGACTGCAAAATTATAT 218
               ||||| |||    ||    | || ||| |   |||  || |  | |  || ||  |||
Db     16664 CATATTAATAT--AATAAATATATTAATTATATAATAAATATATTAATATAATATATTAT 16721

Qy       219 ATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGCATTACATCT 278
             || | ||   | |   | || | ||  ||| ||  || | |||||    |  | ||
Db     16722 ATAATAA-ATATATTAATTAATATAATATATTATATATGTCTTAAATATATAACATATAA 16780
```

```
Qy        279 AAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTC 338
Db      16781 TACATTACATATTATATATAATTATATTATATAATTATAGATAATACATATATTATCTAT 16840

Qy        339 CGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTT 398
Db      16841 AATATATAATATATTATATATTTATATATAAATACATTTATATATTATATATGATAGATA 16900

Qy        399 TACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTT 458
Db      16901 TATATAAATATATTTGTATATTTATATATGATAGATATATAT--------AAATATATTT 16952

Qy        459 TGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTAT 518
Db      16953 GTATATTTATATATGATAGATATATATAAATATATTTGTATATTTATATATGATAGATAT 17012

Qy        519 GCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACTAAT 578
Db      17013 A----TATTTGTATATTTATATATGATAGATATATATTTATATATTTATATATTATAGAT 17068

Qy        579 AGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACATAAA 638
Db      17069 ATATATAAATTTATATATTTATATATTATAGATATATATAAATATATTTATATATTTATA 17128

Qy        639 GCTTTATCATGTTAAAACAACAAAATCTCATTATATA------GCAATTTTACGTAGATC 692
Db      17129 TATTATAGATATATAAATATATTTATATATTTATATATTATAGATATATATAAATATATT 17188

Qy        693 TTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGT 752
Db      17189 TATATATTTATATATTATAGATATATATAAATATATTTATATATTTATATATTATAGATA 17248

Qy        753 CATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTT 812
Db      17249 TATAAATATATTTATATATTTATATATTATAGATATATATAAATATATTTATATATTTAT 17308

Qy        813 TGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGA 872
Db      17309 ATTTATTGATATATATAAATATATTTATATATTTATATATTATAGATATATATAAATATA 17368

Qy        873 ATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCT 932
Db      17369 TTTATAGATTTATATATTATAGATATATATAAATATATTTATAGATTTATATATTATAGA 17428

Qy        933 TGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATT 992
Db      17429 TATATATAAATATATTTATAGATTTATATATTATAGATATATATAAATATATTTATAGAT 17488

Qy        993 TTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGG 1052
Db      17489 TTATATATTATAGATATATATAAA-TATATTTATAGATTTATATATTATAGATATATATA 17547

Qy       1053 CATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAA 1112
Db      17548 AATATATTTATAGATTTATATATTATAGATATATATAAATATATTTATATATATTATAGA 17607

Qy       1113 TTATTTAAAAATACTTTGGATATTAGTA 1140
Db      17608 TATATATAAATATATTTATATATTTATA 17635


RESULT 2
US-10-240-708-73
; Sequence 73, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/240,708
;  CURRENT FILING DATE:  2002-10-03
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 73
;   LENGTH: 5152
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-73

  Query Match             4.7%;  Score 60.4;  DB 10;  Length 5152;
  Best Local Similarity   44.7%;  Pred. No. 1.3;
  Matches  425;  Conservative    0;  Mismatches  501;  Indels   24;  Gaps    4;

Qy        210 AAATTATATATCAGAATGCAGAGCCAAAAACAAAAGTTATGATGCATATATTAAATCAGC 269
Db       2426 AAATAAATTTTAATAATTTAAAAAAAAAAAATAAAATTTTTGTTTTTTATTGGTTTTAGA 2485

Qy        270 ATTACATCTAAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGG 329
Db       2486 ATAAATATAAAAATAGGAAGTTTGTGTTATTTTTGGGTTTTTGTAGTTGTTATTTTTTTT 2545

Qy        330 GGTAATTTCCGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAAT 389
Db       2546 AGTTAGATGGTTTTTTATTATATTTTTTTTTTATTACGTTTTTGTTTAAATATTAAATAT 2605

Qy        390 GCTCTTTTTTACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAA 449
Db       2606 TATTTTTTTGAATAGGTTTTTTCGGTGTAGTTTTAGTTGTTTAAAATTTATAATTTGTTT 2665

Qy        450 AATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATAT 509
Db       2666 TATTTTTTTTGTTTTTATTTGAAATTATAAATTTTGTTTATGGTTTGTTTTTATTAGAAT 2725

Qy        510 GATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGTTGACATAAAATT 569
Db       2726 GTAAGTTTTATAATGATAGGGATATTGTTTTGTTGAATAGTTTTTATAGTATATAGAATA 2785

Qy        570 ATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATATTGTATATAG 629
Db       2786 ATGTGGTATTATAGTAGAGGTTTAAAAATTTGTGTTTTAAAGAAGGTTAATAGTGAAAAA 2845

Qy        630 ATACATAAAGCTTTATCATGTTAAAACAA--------CAAAATCTCATTATATAGCAATT 681
Db       2846 ATGATTTTAATTTTTTTAAGTTTTGAAAATTTTGTTTTAAAATGTAAATTTATGTTTTGT 2905

Qy        682 TTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACA 741
Db       2906 GTTTTTAGAGGTTAAAAAAAATATTTAAGTGTAGTATTTTAAATAATATTTTGAAAATGT 2965

Qy        742 GATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACG 801
Db       2966 AATTTATTTAGTTTGTAATATTTTAATAGTTAGTTTTATATAAATTATTTTTAATATTGG 3025

Qy        802 ATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTT--TGCAAGCATTATGGC 859
Db       3026 TAAAATATTTTTTATTATGTGGGTTTTATATATATATATGTTTAATGGATTTATTTTATT 3085

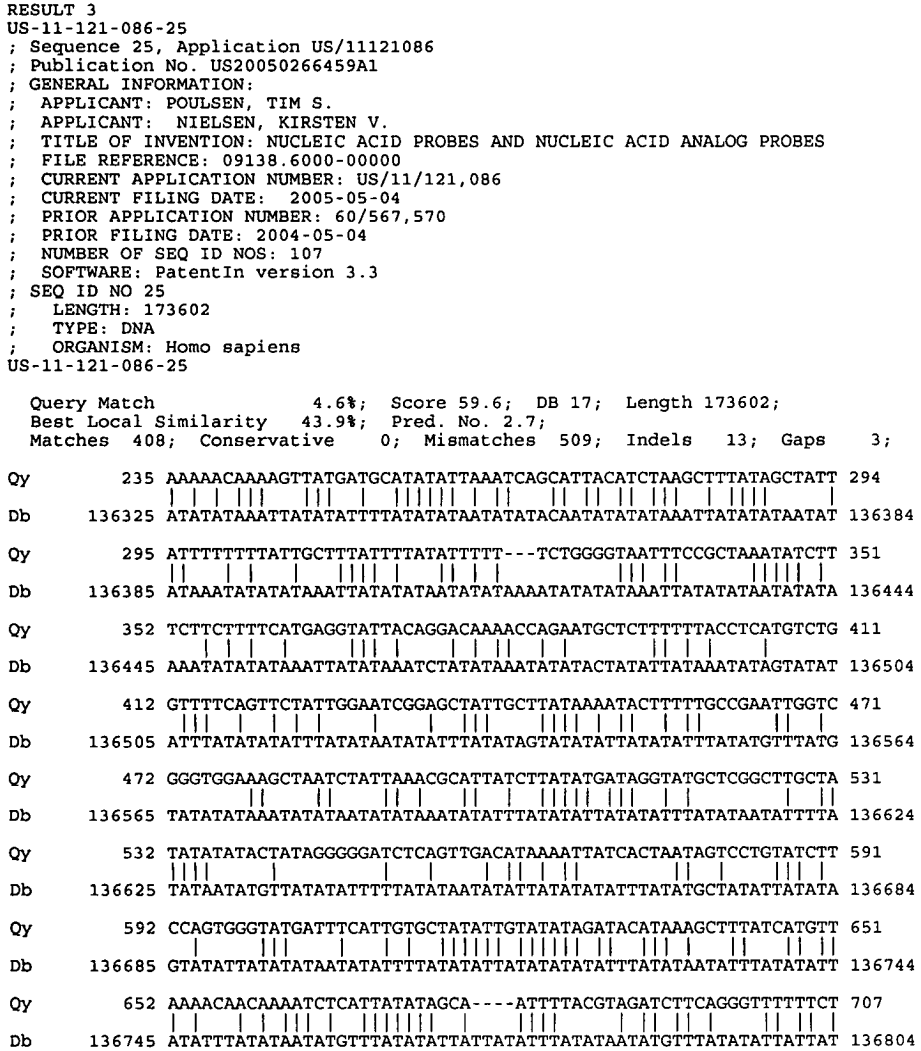
Qy        860 CTATATGTGCTGA----------ATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAA 909
Db       3086 TTATATATGTTGATTGATAATTTTTTGAGTATTTATTATGGATTTAGTATAATATTAGAT 3145

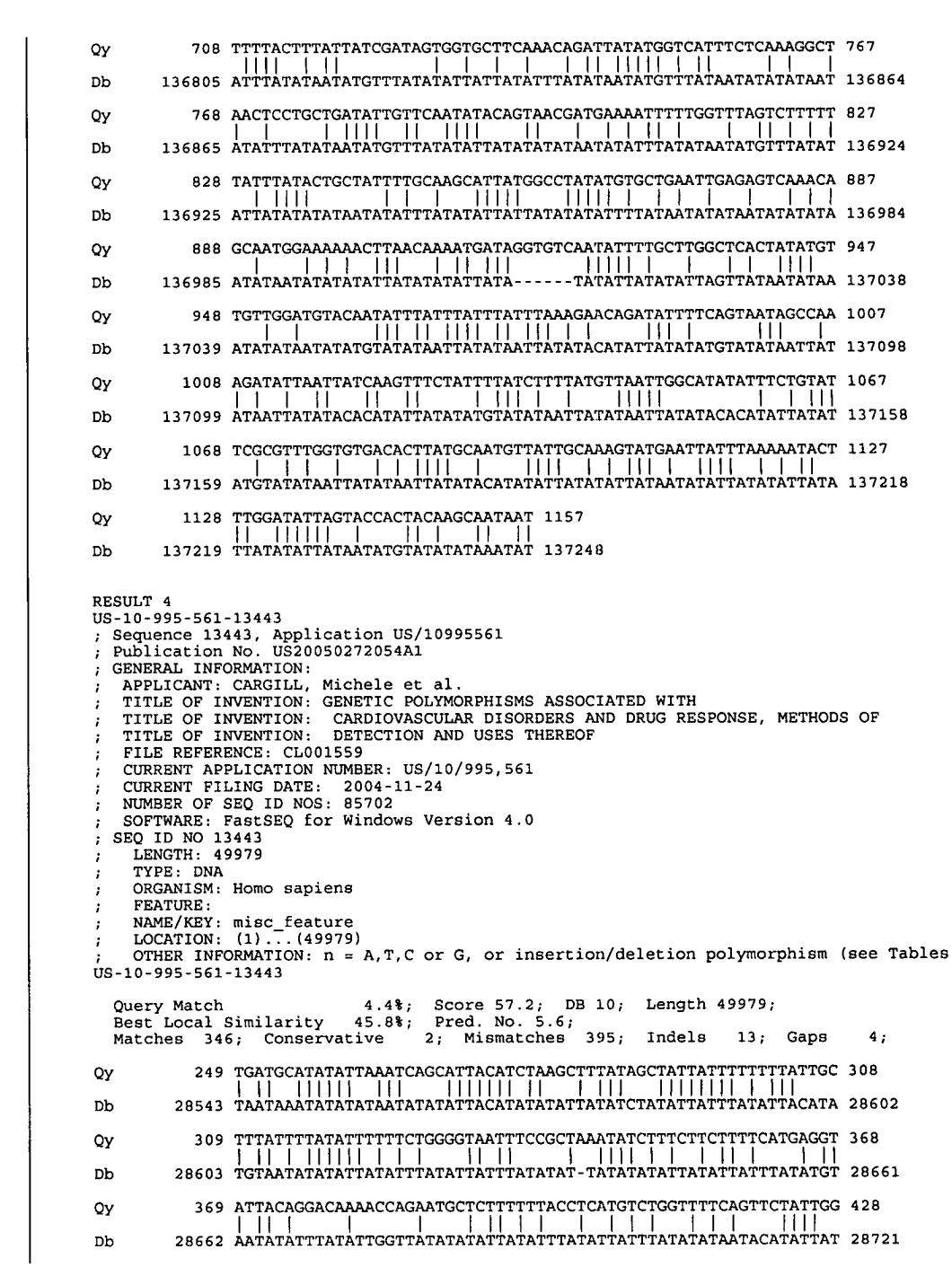
Qy        910 ATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATT 969
```



RESULT 3
US-11-121-086-25
; Sequence 25, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; LENGTH: 173602
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-25

Query Match 4.6%; Score 59.6; DB 17; Length 173602;
Best Local Similarity 43.9%; Pred. No. 2.7;
Matches 408; Conservative 0; Mismatches 509; Indels 13; Gaps 3;





RESULT 4
US-10-995-561-13443
; Sequence 13443, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13443
; LENGTH: 49979
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(49979)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-995-561-13443

Query Match 4.4%; Score 57.2; DB 10; Length 49979;
Best Local Similarity 45.8%; Pred. No. 5.6;
Matches 346; Conservative 2; Mismatches 395; Indels 13; Gaps 4;

```
Qy        429 AATCGGAGCTATTGCTTATAAAATACTTTTTGCCGA-ATTGGTCGGGTGGAAAGCTAATC 487
Db      28722 ATTTATATTATTTATATATAATATATTATTATTTATATTATTTATATATAATATATATA 28781

Qy        488 TATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGG 547
Db      28782 ATAAATATATAATATATATAATAATATATAATTAATAAATATATATAATATATATTACAT 28841

Qy        548 GGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTT 607
Db      28842 ATATATTATATCTATATTATTTATATTACATATGTAATATATATTATATTTATATTATTT 28901

Qy        608 CATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTC 667
Db      28902 ATATAATATATTATATATATTATATTATTTATATGTAATATATTTATATTGKTTATATAT 28961

Qy        668 ATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAG 727
Db      28962 ATTATATTTATATTATTTATATATAATACATATTATATTTATATTATTTATATATAATAT 29021

Qy        728 TGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTC 787
Db      29022 ATATAATAAATATATAATATATATAAAAATATATATATTTAATATATCTATAATATATAT 29081

Qy        788 AATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGC 847
Db      29082 TATATATATTATATATAWWATATATAATTGTACATATATTTATTATATATATTAT----- 29136

Qy        848 AAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACA 907
Db      29137 -----ATATATAATATATATTATAAATATAATATATAAATATATTTATAAATATATATAA 29191

Qy        908 AAATGATAGGTGT-CAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTT 966
Db      29192 ATATTATATTTATACATTATATTTATATACATATTATATATACATTTTTATATATTATAT 29251

Qy        967 ATTTATTTATTTAAAGAACAGATATTTTCAGTAATA 1002
Db      29252 ATATATAAATATAATATACATATATTTTTTAAAATA 29287
```

RESULT 5
US-10-301-480-568791/c
; Sequence 568791, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 568791
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-568791

Query Match 4.4%; Score 56.4; DB 12; Length 990;
Best Local Similarity 44.4%; Pred. No. 4.8;
Matches 269; Conservative 0; Mismatches 336; Indels 1; Gaps 1;

```
Qy        438 TATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGC 497
Db        937 TATATATTATATACTATATATTTTATATTTTATAATATATAATTTATATATATAAATTAC 878

Qy        498 ATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGT 557
Db        877 ATTTTAAAATATAATATATATAAATTATATATAATATATAAATTATTTATATATATTATT 818

Qy        558 TGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTA 617
Db        817 AATATATCATATATTATATATAATATATTACATATAACATAATACAATAT-ATAATCATA 759

Qy        618 TATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGC 677
Db        758 TATTATATTACATTATATTATATATTATATTTTAAATAAATTATATATTATAAAATAATA 699

Qy        678 AATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCA 737
Db        698 TAATTATGTTACATTATAAAATAATATATTATTACATTATATAACATTATTATATTATAT 639

Qy        738 AACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGT 797
Db        638 TATATATAATATATTTAATACAAACTATTATGTTATATTTAATATAATATATAATATAAT 579

Qy        798 AACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATG 857
Db        578 ATTATATTTTATATTTAATATAATATATAACATATTTTATATTTAATATAAAATATAATA 519

Qy        858 GCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGG 917
Db        518 TATATTATATTTTATATTTAATATATTATATTTTATATTATTTTATATTTAATATAAT 459

Qy        918 TGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATT 977
Db        458 ATATATTATATTTTATATTATTTTATATTTAATATAATATATTATTATATTATTTTATA 399

Qy        978 TAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATC 1037
Db        398 TTATATTACTATATTATATATTATAGAATATATTATAGAATAGATACAATATATTACATA 339

Qy       1038 TTTTAT 1043
Db        338 TATAAT 333
```

RESULT 6
US-10-301-480-1182200/c
; Sequence 1182200, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1182200
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-1182200

Query Match 4.4%; Score 56.4; DB 12; Length 990;
Best Local Similarity 44.4%; Pred. No. 4.8;
Matches 269; Conservative 0; Mismatches 336; Indels 1; Gaps 1;

```
Qy        438 TATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGC 497
Db        937 TATATATTATATACTATATATTTTATATTTTATAATATATAATTTATATATATAAATTAC 878

Qy        498 ATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACTATAGGGGGATCTCAGT 557
```

```
Db        877 ATTTTAAAATATAATATATATAAATTATATATAATATATAAATTATTTATATATATTATT 818

Qy        558 TGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTA 617
              |  ||| |  ||| |   ||  |  |   ||  || | ||  |  ||
Db        817 AATATATCATATATTATATATAATATTACATATAACATAATACAATAT-ATAATCATA 759

Qy        618 TATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGC 677
              |||| |||     || || |    ||||  | | || | |  | |   |    |   |
Db        758 TATTATATTACATTATATTATATATTATATTTTAAATAAATTATATATTATAAAATAATA 699

Qy        678 AATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCA 737
              | ||    || ||  | |     | |  | ||   ||  || |  || |      |
Db        698 TAATTATGTTACATTATAAAATAATATATTATTACATTATATAACATTATTATATTATAT 639

Qy        738 AACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGT 797
              | | || ||||  | | | |  |       |     | ||||  |  |  ||| | |
Db        638 TATATATAATATATTTAATACAAACTATTATGTTATATTTAATATAATATATAATAATAT 579

Qy        798 AACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATG 857
              |        || |||  | || | | |   || | |  |  |  |   ||  || ||
Db        578 ATTATATTTTATATTTAATATAATATATAACATATTTTATATTTAATATAAAATATAATA 519

Qy        858 GCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGG 917
                 ||| |  |  ||| |   |  |  |     |   | |  |||      |  |||
Db        518 TATATTATATTTTATATTTAATATATATTATATTTTATATTATTTTATATTTAATATAAT 459

Qy        918 TGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATT 977
                | ||| ||   |     |||||  |   |   || || | ||| |||| |||||
Db        458 ATATATTATATTTTATATTATTTTATATTTAATATAATATATATTATATATTATTTTATA 399

Qy        978 TAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATC 1037
              | |       ||||| |   | ||||   |   |||  | ||   |   | ||||  ||
Db        398 TTATATTACTATATTATATATTATAGAATATATTATAGAATAGATACAATATATTACATA 339

Qy       1038 TTTTAT 1043
              | | ||
Db        338 TATAAT 333


RESULT 7
US-10-240-708-26
; Sequence 26, Application US/10240708
; Publication No. US20050282157A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/240,708
;  CURRENT FILING DATE:  2002-10-03
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 26
;   LENGTH: 6583
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-708-26

  Query Match             4.3%;  Score 55.2;  DB 10;  Length 6583;
  Best Local Similarity   43.6%;  Pred. No. 9.3;

  Matches  346;  Conservative    0;  Mismatches  443;  Indels    5;  Gaps    2;

Qy        279 AAGCTTTATAGCTATTATTTTTTTTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTC 338
              ||  ||| ||  |||||||  |   ||  |  |  |  |  |    | || | |    | || |
Db       5779 AATTTTTTTATTTATTATTATAAATAGTAATATTGTATTATTATTATATTATATTATA 5838

Qy        339 CGCTAAATATCTTTCTTCTTTTCATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTT 398
              |  |||||| | | |        || |  | | |    |  |    | ||||    | |
Db       5839 TGTAAAATATATATATAGAAAGTATAAATTTTAAGTAAATTATTATAAAATGAATATATA 5898

Qy        399 TACCTCATGTCTGGTTTTCAGTTCTATTGGAATCGGAGCTATTGCTTATAAAATACTTTT 458
              |   |  | |||||||    |   |       ||| || |  |  |||   |    ||||
Db       5899 TTGTTTTTATTTGGTTTTTTTTAGAAGATATTTCGTAGTTCGTTTTTAATTAGATATTTT 5958

Qy        459 TGCCGAATTGGTCGGGTGGAAAGCTAATCTATTAAACGCATTATCTTATATGATAGGTAT 518
              ||     |  |    | |     | ||             ||   || |   ||   |||
Db       5959 TGTTTTAATTATTAATTTGTTTTTTTATTGTAAGTGTTGTTTCGTTTGTGAAATGTATAT 6018

Qy        519 GCTCGGCTTGCTATATATATACTATAGGGGGAT----CTCAGTTGACATAAAATTATCAC 574
                  |  || |  ||  | |  | |    | ||      |   || |  ||| |||||
Db       6019 AAAAGAAATGATGTAGTTTTTATTTTTTAGTATTTGGTTTTTTTTATTTAATATTATGTT 6078

Qy        575 TAATAGTCCTGTATCTTCCAGTGGGTATGATTTCATTGTGCTATAT-TGTATATAGATAC 633
              |   ||   | | | |   | | | |  ||  | | ||  | | | | | | |||    |
Db       6079 TGTGAGATTTATTTGTATTATCGAGAGTATTTATAGTTTGTTTTTTGTTTGTATTTTAAT 6138

Qy        634 ATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCT 693
              || |  ||||| |||| |  | |  | ||| |  |||| |     ||     |    |
Db       6139 GTAGAATTTTATTATGTAAGTATATTATAATTTATTTATTTTATTATTGATAGATATTTG 6198

Qy        694 TCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTC 753
                 || |||    ||||    ||||| |    |      |  ||     |  |||| ||
Db       6199 AGTGGTTTTAAGATTTTATTATTATTGTTTTTGTGAATATTTTTGTAATTTTATGTGTAT 6258

Qy        754 ATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTT 813
              |   |  ||   | | |  |  || | ||||| | | ||   ||   ||  |  |||
Db       6259 GTGATTTTAATTTTTATTTTTTTTGTTTTGTTTATTTTATTTTAGTTATTTAGTTTTCG 6318

Qy        814 GGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAA 873
              |||   | ||   ||| |    |   ||||      |    |    |||| ||  |
Db       6319 TGTTATTTTTTGAATATGTTAGGTATATTTTTATTTTTAGGGTTTTGTATACGTTGTTTT 6378

Qy        874 TTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTT 933
              ||     |  ||       |  || | |              |  | | | |  || |  |
Db       6379 TTTTTTTTAGAATGTTTTTTTAATAGATAGTTTCGTAGTTTGGTTTTTTATTTCGTTTAG 6438

Qy        934 GGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTT 993
                |   |||| ||| ||    |     |    || |||| |||| | |   | |  | |
Db       6439 TAATATATATAGGTTTTTTTTTAGTGAGATATTTTTTAATTATGTTATTTAAAATTGTAA 6498

Qy        994 TCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGC 1053
              |   || |          || |||     ||| | ||| || | || | ||   ||
Db       6499 TTTTTATTTTTATTTTGGGATATTTAATTTTTTTTTTTTTGATTTATTTTTTTTTATGAT 6558

Qy       1054 ATATATTTCTGTAT 1067
              ||||||   || ||
Db       6559 ATATATAATTGGAT 6572


RESULT 8
US-11-121-086-105/c
; Sequence 105, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
;  APPLICANT: POULSEN, TIM S.
;  APPLICANT:  NIELSEN, KIRSTEN V.
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;  FILE REFERENCE: 09138.6000-00000
;  CURRENT APPLICATION NUMBER: US/11/121,086
;  CURRENT FILING DATE:  2005-05-04
;  PRIOR APPLICATION NUMBER: 60/567,570
;  PRIOR FILING DATE: 2004-05-04
;  NUMBER OF SEQ ID NOS: 107
```

```
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 105
;   LENGTH: 171486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-11-121-086-105

  Query Match             4.2%;  Score 54.2;  DB 17;  Length 171486;
  Best Local Similarity   46.6%;  Pred. No. 20;
  Matches  281;  Conservative    0;  Mismatches  313;  Indels    9;  Gaps    3;

Qy        533 ATATATACTATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTC 592
              || |||| ||||     || | | |  |  || |  | ||  |||||      ||| |
Db      17682 ATCTATAATATATAAATATATAAATATATTTATATATCTCTATAATATATAAATATATAA 17623

Qy        593 CAGTGGGTATGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTA 652
                |  |||  |  |           | |||||  || ||| | || ||  || | |
Db      17622 ATATATTTATATATATCTATAATATATATAAATATATTTATATATATCTATAATATATAA 17563

Qy        653 AAACAACAAAATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTA 712
              |   | || || |  ||||| | ||  ||  || |||| |    | ||| | | ||
Db      17562 ATCTATAAATATATTTATATATATCTATAATATATAAATCTATAAATATATTTATATATA 17503

Qy        713 -CTTTATTATCGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACT 771
               || || |||  | |    |  |  |     |||||    | |    |||      |
Db      17502 TCTATAATATATAAATCTATAAATATATTTATATATATCTATAATATATAAATCTATAAA 17443

Qy        772 CCTGCTGATATTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATT 831
                |  |||| || |||     |||  |  |||| | |   |  | | | | ||
Db      17442 TATATTTATATATATCTATAATATATAAATCTATAAATATATTTATATATATCTATAATA 17383

Qy        832 TATACTGCTATTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAA 891
              ||||   ||||      ||  ||  |     ||||| |      ||  | |   | | |   |
Db      17382 TATAAATCTAT-----AAATATATTTATATATATCTATAATATATAAATATATAAATATA 17328

Qy        892 TGGAAAAAACTTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTT 951
              |  | |  |  | || |||   | |  | | ||||| |   ||   |   | |||  | |
Db      17327 TTTATATATATCAATAAATATAAATATATAAATATAT---TTATATATATCTATAATATA 17271

Qy        952 GGATGTACAATATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGAT 1011
                |      | || ||||||| ||| || |  | | | | |  |   |||   | | ||
Db      17270 TAAATATATAAATATATTTATATATCTATAATATATAAATATATAAATATATTTATATAT 17211

Qy       1012 ATTAATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGC 1071
              ||  || ||  |       |  || | ||    ||  |     |||| |   |||
Db      17210 ATCTATAATATATAAATATATAAATATATTTATATATATCTATAATATATAAATATATAA 17151

Qy       1072 GTTTGGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGG 1131
              | |  |   | |  |||   || | | |  | |    | |||| || |  ||   |
Db      17150 ATATATTTATATATCTATAATATATAAATATATAAATATATTTATATATATCTATAATAT 17091

Qy       1132 ATA 1134
              |||
Db      17090 ATA 17088


RESULT 9
US-11-011-332A-77
; Sequence 77, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
;  APPLICANT: Foekens, John
;  APPLICANT:  Harbeck, Nadia
;  APPLICANT:  Koenig, Thomas
;  APPLICANT:  Maier, Sabine
;  APPLICANT:  Martens, John
;  APPLICANT:  Model, Fabian
;  APPLICANT:  Nimmrich, Inko
;  APPLICANT:  Rujan, Tamas
;  APPLICANT:  Schmitt, Manfred
;  APPLICANT:  Lesche, Ralf
;  APPLICANT:  Dietrich, Dimo
;  APPLICANT:  Mueller, Volkmar
;  APPLICANT:  Kluth, Antje
;  APPLICANT:  Schwope, Ina
;  APPLICANT:  Hartmann, Oliver
;  APPLICANT:  Adorjan, Peter
;  TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SUI
;  TITLE OF INVENTION:  BREAST CELL PROLIFERATIVE DISORDER PATIENTS
;  FILE REFERENCE: 47675-99
;  CURRENT APPLICATION NUMBER: US/11/011,332A
;  CURRENT FILING DATE:  2004-12-13
;  PRIOR APPLICATION NUMBER: US 10/517,741
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: PCT/EP2003/010881
;  PRIOR FILING DATE: 2003-10-01
;  PRIOR APPLICATION NUMBER: DE 10245779.4
;  PRIOR FILING DATE: 2002-10-01
;  PRIOR APPLICATION NUMBER: DE 10300096.8
;  PRIOR FILING DATE: 2003-01-07
;  PRIOR APPLICATION NUMBER: DE 10317955.0
;  PRIOR FILING DATE: 2003-04-17
;  PRIOR APPLICATION NUMBER: PCT/EP2004/014170
;  PRIOR FILING DATE: 2004-12-13
;  PRIOR APPLICATION NUMBER: EP 03090432.0
;  PRIOR FILING DATE: 2003-12-11
;  PRIOR APPLICATION NUMBER: EP 04090041.7
;  PRIOR FILING DATE: 2004-02-10
;  PRIOR APPLICATION NUMBER: EP 04090380.9
;  PRIOR FILING DATE: 2004-09-30
;  PRIOR APPLICATION NUMBER: EP 04090127.4
;  PRIOR FILING DATE: 2004-04-01
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 77
;   LENGTH: 6499
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-77

  Query Match             4.2%;  Score 54;  DB 17;  Length 6499;
  Best Local Similarity   42.4%;  Pred. No. 14;
  Matches  427;  Conservative    0;  Mismatches  565;  Indels   14;  Gaps    2;

Qy        242 AAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTT 301
              || ||||| ||   ||| || | |      ||  || |      || | |  || | ||
Db       4330 AAGGTTATAATATTTATTTTTATTTGAATTTTTATTTTGTTAGTAAATTTTATAGATGTA 4389

Qy        302 TTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTC 361
                |||  |  |||||||| ||  |    |  |||     |  | || ||| || ||||
Db       4390 GGATTTGTAGTTTTTATAATTATGTGAGTTAATTTTTTAATGTAAATTTTTTTTATTTTT 4449

Qy        362 ATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTT 421
              |   ||  ||       |    |  | ||| ||   ||   |  || |   ||
Db       4450 TTTTGTTATATTTATAGATTTATTTGTTATATTTATTTATGTATTATTTATTATTTATT 4509

Qy        422 CTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAG 481
              | ||   | |      ||||  |||| |  |   |||       |   |       |
Db       4510 ATTTATTATTTATTTATATTATTTATTATTTTTATTTATTTATTTATTTATTTATT 4569

Qy        482 CTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACT 541
                | | |        |||   |||     ||   | |  | || ||  | | | ||||
Db       4570 TATTTATTTATTTATTATTTATTTAATATATTTATTTTATTGGTTTTGTTTTTTTATATA 4629

Qy        542 ATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTA 601
              ||  |     | | ||| |  ||  |||  |    |  |  | | | ||        |||
Db       4630 ATTTTGATTAATAAATTTATTAAATTTTAAGAAGTTTTTTTATTTTTTTTAGGTGTAGTA 4689

Qy        602 TGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAA 661
              |   ||  || ||  ||||    | |    | |    |  | |  ||||   |  | |
Db       4690 TATATTTTTTTTGTTATAGGTGAGAAGTGTTAAGGTTGTAATTTGTTGTTTTTATTAGAG 4749

Qy        662 AATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTAT 721
                ||  |||   |  | ||| ||  |   | ||  |  | | |   | |    || ||| |
```



RESULT 10
US-10-301-480-602391/c
; Sequence 602391, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 602391
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-602391

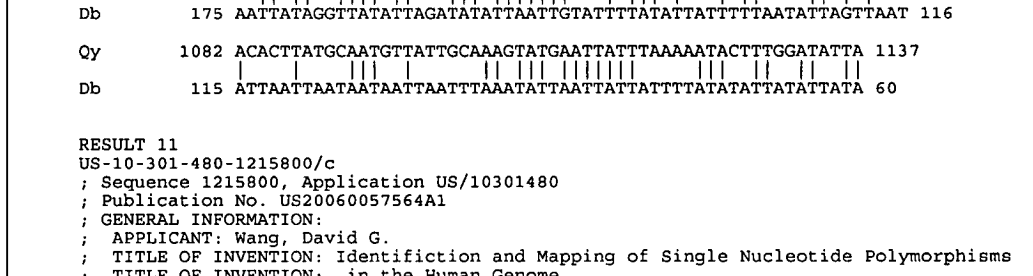
Query Match 4.1%; Score 53.6; DB 12; Length 607;
Best Local Similarity 46.3%; Pred. No. 13;
Matches 248; Conservative 0; Mismatches 284; Indels 4; Gaps 2;





RESULT 11
US-10-301-480-1215800/c
; Sequence 1215800, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1215800
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-1215800

Query Match 4.1%; Score 53.6; DB 12; Length 607;
Best Local Similarity 46.3%; Pred. No. 13;
Matches 248; Conservative 0; Mismatches 284; Indels 4; Gaps 2;





```
Qy        846 GCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAA 905
Db        355 GGTAATAATTATATATAATTATATATTATATTATATGTAATAATTATATATAAGTATATA 296

Qy        906 CAAAATGATAGGTGTCAAT--ATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATA 963
Db        295 TTATATTATATGTAATAATTATATATAATTTATGTATTATATTATATGTAATAAATACAC 236

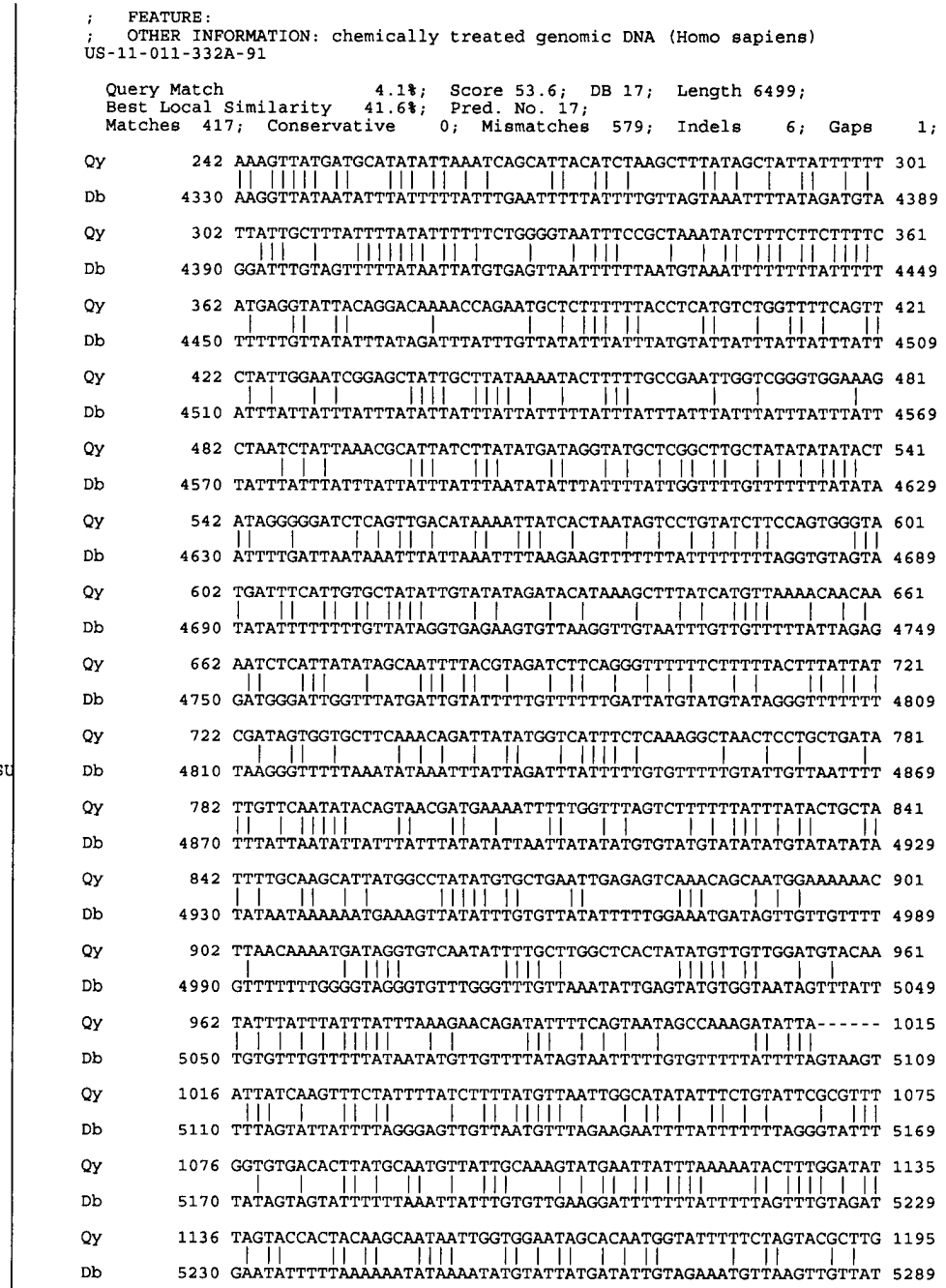
Qy        964 TTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAA 1023
Db        235 ATAATTATATATATTTTATTATATATTATTGTATAATTGTATATTATATTATATATAAT 176

Qy       1024 GTTTCTATTTTATCTT--TTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTG 1081
Db        175 AATTATAGGTTATATTAGATATATTAATTGTATTTTATATTATTTTTAATATTAGTTAAT 116

Qy       1082 ACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTA 1137
Db        115 ATTAATTAATAATAATTAATTTAAATATTAATTATTATTTTATATATTATATTATA 60
```

RESULT 12
US-11-011-332A-91
; Sequence 91, Application US/11011332A
; Publication No. US20060024684A1
; GENERAL INFORMATION:
; APPLICANT: Foekens, John
; APPLICANT: Harbeck, Nadia
; APPLICANT: Koenig, Thomas
; APPLICANT: Maier, Sabine
; APPLICANT: Martens, John
; APPLICANT: Model, Fabian
; APPLICANT: Nimmrich, Inko
; APPLICANT: Rujan, Tamas
; APPLICANT: Schmitt, Manfred
; APPLICANT: Lesche, Ralf
; APPLICANT: Dietrich, Dimo
; APPLICANT: Mueller, Volkmar
; APPLICANT: Kluth, Antje
; APPLICANT: Schwope, Ina
; APPLICANT: Hartmann, Oliver
; APPLICANT: Adorjan, Peter
; TITLE OF INVENTION: PROGNOSTIC MARKERS FOR PREDICTION OF TREATMENT RESPONSE AND/OR SU
; TITLE OF INVENTION: BREAST CELL PROLIFERATIVE DISORDER PATIENTS
; FILE REFERENCE: 47675-99
; CURRENT APPLICATION NUMBER: US/11/011,332A
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US 10/517,741
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: PCT/EP2003/010881
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: DE 10245779.4
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: DE 10300096.8
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: DE 10317955.0
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: PCT/EP2004/014170
; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: EP 03090432.0
; PRIOR FILING DATE: 2003-12-11
; PRIOR APPLICATION NUMBER: EP 04090041.7
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: EP 04090380.9
; PRIOR FILING DATE: 2004-09-30
; PRIOR APPLICATION NUMBER: EP 04090127.4
; PRIOR FILING DATE: 2004-04-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 158
; SEQ ID NO 91
; LENGTH: 6499
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-11-011-332A-91

Query Match 4.1%; Score 53.6; DB 17; Length 6499;
Best Local Similarity 41.6%; Pred. No. 17;
Matches 417; Conservative 0; Mismatches 579; Indels 6; Gaps 1;



```
Qy        242 AAAGTTATGATGCATATATTAAATCAGCATTACATCTAAGCTTTATAGCTATTATTTTTT 301
Db       4330 AAGGTTATAATATTTATTTTTATTTGAATTTTTATTTTGTTAGTAAATTTTATAGATGTA 4389

Qy        302 TTATTGCTTTATTTTATATTTTTTCTGGGGTAATTTCCGCTAAATATCTTTCTTCTTTTC 361
Db       4390 GGATTTGTAGTTTTTATAATTATGTGAGTTAATTTTTTAATGTAAATTTTTTTTATTTTT 4449

Qy        362 ATGAGGTATTACAGGACAAAACCAGAATGCTCTTTTTTACCTCATGTCTGGTTTTCAGTT 421
Db       4450 TTTTTGTTATATTTATAGATTTATTTGTTATATTTATTTATGTATTATTTATTATTTATT 4509

Qy        422 CTATTGGAATCGGAGCTATTGCTTATAAAATACTTTTTGCCGAATTGGTCGGGTGGAAAG 481
Db       4510 ATTTATTATTTATTTATATTATTTATTATTTTTATTTATTTATTTATTTATTTATTT 4569

Qy        482 CTAATCTATTAAACGCATTATCTTATATGATAGGTATGCTCGGCTTGCTATATATATACT 541
Db       4570 TATTTATTTATTTATTATTTATTTAATATATTTATTTTATTGGTTTTGTTTTTTTATATA 4629

Qy        542 ATAGGGGGATCTCAGTTGACATAAAATTATCACTAATAGTCCTGTATCTTCCAGTGGGTA 601
Db       4630 ATTTTGATTAATAAATTTATTAAATTTTAAGAAGTTTTTTTATTTTTTTTAGGTGTAGTA 4689

Qy        602 TGATTTCATTGTGCTATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAA 661
Db       4690 TATATTTTTTTTGTTATAGGTGAGAAGTGTTAAGGTTGTAATTTGTTGTTTTTATTAGAG 4749

Qy        662 AATCTCATTATATAGCAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTAT 721
Db       4750 GATGGGATTGGTTTATGATTGTATTTTTGTTTTTTGATTATGTATGTATAGGGTTTTTTT 4809

Qy        722 CGATAGTGGTGCTTCAAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATA 781
Db       4810 TAAGGGTTTTTAAATATAAATTTATTAGATTTATTTTTGTGTTTTTGTATTGTTAATTTT 4869

Qy        782 TTGTTCAATATACAGTAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTA 841
Db       4870 TTTATTAATATTATTTATTTATATATTAATTATATATGTGTATGTATATATGTATATATA 4929

Qy        842 TTTTGCAAGCATTATGGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAAC 901
Db       4930 TATAATAAAAAATGAAAGTTATATTTGTGTTATATTTTTGGAAATGATAGTTGTTGTTTT 4989

Qy        902 TTAACAAAATGATAGGTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAA 961
Db       4990 GTTTTTTTGGGGTAGGGTGTTTGGGTTTGTTAAATATTGAGTATGTGGTAATAGTTTATT 5049

Qy        962 TATTTATTTATTTATTTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTA------ 1015
Db       5050 TGTGTTTGTTTTTATAATATGTTGTTTTATAGTAATTTTTGTGTTTTTATTTTAGTAAGT 5109

Qy       1016 ATTATCAAGTTTCTATTTTATCTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTT 1075
Db       5110 TTTAGTATTATTTTAGGGAGTTGTTAATGTTTAGAAGAATTTTATTTTTTTAGGGTATTT 5169

Qy       1076 GGTGTGACACTTATGCAATGTTATTGCAAAGTATGAATTATTTAAAAATACTTTGGATAT 1135
Db       5170 TATAGTAGTATTTTTTAAATTATTTGTGTTGAAGGATTTTTTTATTTTTAGTTTGTAGAT 5229

Qy       1136 TAGTACCACTACAAGCAATAATTGGTGGAATAGCACAATGGTATTTTTCTAGTACGCTTG 1195
Db       5230 GAATATTTTTAAAAAATATAAAATATGTATTATGATATTGTAGAAATGTTAAGTTGTTAT 5289

Qy       1196 GAATCAGTGGAGTGCTGCTTGGCTTGATTATATCTTTTGCTT 1237
```

```
                ||    |  |  |  | ||    |  || |||  |||  ||
Db       5290 AAAAGTTTAAAAGTAGATTTTTATTTTTTTATTATTTTTTT 5331
```

RESULT 13
US-10-301-480-568793/c
; Sequence 568793, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 568793
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-568793

Query Match 4.1%; Score 53; DB 12; Length 990;
Best Local Similarity 44.6%; Pred. No. 17;
Matches 250; Conservative 0; Mismatches 310; Indels 1; Gaps 1;

```
Qy        617 ATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAG 676
              ||||   |  ||  || ||| |   |||  || | | |       | ||   ||  || |
Db        703 ATATATAAATTATTTATATATATTATTAATATATCATATATTATATATAATATATTACAT 644

Qy        677 CAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTC 736
                |   ||      || |  |     | | |  | |||| |||| |  ||| |  |  ||
Db        643 ATAACATAATACAATATATAATCATATATTATATTACATTATATTATATA-TTATATTTT 585

Qy        737 AAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAG 796
              ||| | |||||||  |       | | |   | |       | |||   |   ||||
Db        584 AAATAAATTATATATTATAAAATAATATAATTATGTTACATTATAAAATAATATATTATT 525

Qy        797 TAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTAT 856
              |  ||  ||  || |  | |  | || | ||| | || | | |  |   | |  |||
Db        524 ACATTATATAACATTATTATATTATATTATATATATTTAATACAAACTATTATGTTATAT 465

Qy        857 GGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAG 916
              ||||     | |  | | | |  |      || || | | |  | || ||  |
Db        464 TTAATATAATATATAATATAATATTATATTTTATATTTAATATAATATATAACATATTTT 405

Qy        917 GTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTAT 976
              | |  |   |    |   |    |  |||   | ||  || ||  ||| |   |  | |||
Db        404 ATATTTAATATAAAATATAATATATATTATATTTTATATTTAATATAATATATATTATAT 345

Qy        977 TTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTAT 1036
              || |          ||| || | ||  |   | |      | |     | ||| ||||| ||
Db        344 TTTATATTATTTTATATTTAATATAATATATATTATATTTTATATTATTTTATATTTAAT 285

Qy       1037 CTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGT 1096
              |  ||| |    |    ||| ||    |||     | |  | |       ||    || |
Db        284 ATAATATATATTATATATTATTTTATATTATATTACTATATTATATATTATAGAATATAT 225

Qy       1097 TATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTACCACTACAAGCAATAA 1156
              ||| | | ||    |||   ||| | |||   |  |  | | |   |    ||  ||| |
Db        224 TATAGAATAGATACAATATATTACATATATAATACAAGTAAATTATATATAAATAAATTA 165

Qy       1157 TTGGTGGAATAGCACAATGGT 1177
              |  ||   ||   |||  |||
Db        164 TATGTAATATGTTACATAGGT 144

RESULT 14

US-10-301-480-1182202/c
; Sequence 1182202, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1182202
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-1182202

Query Match 4.1%; Score 53; DB 12; Length 990;
Best Local Similarity 44.6%; Pred. No. 17;
Matches 250; Conservative 0; Mismatches 310; Indels 1; Gaps 1;

```
Qy   617 ATATTGTATATAGATACATAAAGCTTTATCATGTTAAAACAACAAAATCTCATTATATAG 676
              |||| | || || ||| | ||| || | | | | || || || |
Db   703 ATATATAAATTATTTATATATATTATTAATATATCATATATTATATATAATATATTACAT 644

Qy   677 CAATTTTACGTAGATCTTCAGGGTTTTTTCTTTTTACTTTATTATCGATAGTGGTGCTTC 736
                | || || | | | | | | |||| |||| | ||| | | ||
Db   643 ATAACATAATACAATATATAATCATATATTATATTACATTATATTATATA-TTATATTTT 585

Qy   737 AAACAGATTATATGGTCATTTCTCAAAGGCTAACTCCTGCTGATATTGTTCAATATACAG 796
              ||| | ||||||| | | | | | | | ||| | ||||
Db   584 AAATAAATTATATATTATAAAATAATATAATTATGTTACATTATAAAATAATATATTATT 525

Qy   797 TAACGATGAAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTAT 856
              | || || || | | | | || | ||| | || | | | | | | |||
Db   524 ACATTATATAACATTATTATATTATATTATATATATTTAATACAAACTATTATGTTATAT 465

Qy   857 GGCCTATATGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAG 916
              |||| | | | | | | | || || | | | | || || |
Db   464 TTAATATAATATATAATATAATATTATATTTTATATTTAATATAATATATAACATATTTT 405

Qy   917 GTGTCAATATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTAT 976
              | | | | | | | ||| | || || || ||| | | | |||
Db   404 ATATTTAATATAAAATATAATATATATTATATTTTATATTTAATATAATATATATTATAT 345

Qy   977 TTAAAGAACAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTAT 1036
              || | ||| || | || | | | | | | ||| ||||| ||
Db   344 TTTATATTATTTTATATTTAATATAATATATATTATATTTTATATTATTTTATATTTAAT 285

Qy   1037 CTTTTATGTTAATTGGCATATATTTCTGTATTCGCGTTTGGTGTGACACTTATGCAATGT 1096
              | ||| | | ||| || ||| | | | | || || |
Db   284 ATAATATATATTATATATTATTTTATATTATATTACTATATTATATATTATAGAATATAT 225

Qy   1097 TATTGCAAAGTATGAATTATTTAAAAATACTTTGGATATTAGTACCACTACAAGCAATAA 1156
              ||| | | || ||| ||| | ||| | | | | | | || ||| |
Db   224 TATAGAATAGATACAATATATTACATATATAATACAAGTAAATTATATATAAATAAATTA 165

Qy   1157 TTGGTGGAATAGCACAATGGT 1177
              | || || ||| |||
Db   164 TATGTAATATGTTACATAGGT 144

RESULT 15
US-10-301-480-563203
; Sequence 563203, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.

```
;  TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
;  TITLE OF INVENTION:  in the Human Genome
;  FILE REFERENCE: 108827.137
;  CURRENT APPLICATION NUMBER: US/10/301,480
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 10/215,598
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: US 60/311,695
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 1226818
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 563203
;   LENGTH: 481
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-301-480-563203

  Query Match             4.1%;  Score 52.8;  DB 12;  Length 481;
  Best Local Similarity   53.0%;  Pred. No. 17;
  Matches  141;  Conservative    0;  Mismatches  117;  Indels    8;  Gaps    1;

Qy        805 AAAATTTTTGGTTTAGTCTTTTTTATTTATACTGCTATTTTGCAAGCATTATGGCCTATA 864
              ||||| | |  | ||    | | |  | | |    |||| |  |   |||||    ||||
Db          8 AAAATATATATTATAAATATATATTATAAAATATATATTATAAATATATTATAAAATATA 67

Qy        865 TGTGCTGAATTGAGAGTCAAACAGCAATGGAAAAAACTTAACAAAATGATAGGTGTCAAT 924
              | |  | |||  |   | ||| |   ||   ||| |  | | |||||      | | |||
Db         68 TATTATAAATATATTATAAAATATATATTATAAATATATTATAAAATATATATTATAAAT 127

Qy        925 ATTTTGCTTGGCTCACTATATGTTGTTGGATGTACAATATTTATTTATTTATTTAAAGAA 984
              ||  |   |     | ||||||| |   || || || || |  | |  ||| | | |||
Db        128 ATAATATATTATAAAATATATGTTATAAAATATAAAACATGTTATAAAATATATTATGAA 187

Qy        985 CAGATATTTTCAGTAATAGCCAAAGATATTAATTATCAAGTTTCTATTTTATCTTTTATG 1044
                  |  || |||    || ||||  ||||| || | | | || ||| || ||
Db        188 A--------TATGTTATATTTTAAAATATATATTATGAAATATATGTTATATATTATAAA 239

Qy       1045 TTAATTGGCATATATTTCTGTATTCG 1070
               ||  ||  |||||| | || | | |
Db        240 ATATATGTAATATATATATGCAGTTG 265


Search completed: May 12, 2006, 06:55:12
Job time : 661.783 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 08:31:04 ; Search time 3606.55 Seconds
(without alignments)
15792.702 Million cell updates/sec

Title: US-09-423-093B-4_COPY_14059_15060
Perfect score: 1002
Sequence: 1 atgcttatatcattttgtat.....ctaaacggataattaagtag 1002

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1002	100.0	22080	1	SERFBB	X56793 S.enterica
	2	1002	100.0	22080	6	BD085557	BD085557 Nucleic a
	3	1002	100.0	22080	6	AX029264	AX029264 Sequence
c	4	1000.4	99.8	22043	1	AE008792	AE008792 Salmonell
	5	949.4	94.8	3651	1	AY062284	AY062284 Salmonell
	6	360.4	36.0	110000	1	CP000026_08	Continuation (9 of
	7	358.8	35.8	2879	1	SERFBD1B	M65054 Salmonella
c	8	358.8	35.8	250050	1	AL627273	AL627273 Salmonell
	9	358.8	35.8	300928	1	AE016836	AE016836 Salmonell
	10	280.2	28.0	4115	1	AY062287	AY062287 Salmonell
	11	277.8	27.7	1832	1	AY062288	AY062288 Salmonell
	12	277.8	27.7	1832	1	AY062290	AY062290 Salmonell
	13	277.8	27.7	1832	1	AY062291	AY062291 Salmonell
	14	277.8	27.7	1832	1	AY062292	AY062292 Salmonell
	15	277.8	27.7	1834	1	AY062289	AY062289 Salmonell
	16	276.2	27.6	1910	1	AY062285	AY062285 Salmonell
	17	276.2	27.6	1910	1	AY062286	AY062286 Salmonell
	18	269.2	26.9	783	1	AY062293	AY062293 Salmonell
	19	247.6	24.7	26913	1	YEU46859	U46859 Yersinia en
	20	221	22.1	19457	1	AF461769	AF461769 Yersinia
	21	221	22.1	20125	1	AF461770	AF461770 Yersinia
	22	129.4	12.9	3742	1	SCU04165	U04165 Salmonella
	23	62.6	6.2	3474	2	PFU85963	U85963 Plasmodium
c	24	62.6	6.2	4300	2	PFSC04095	AL010267 Plasmodiu
	25	62.6	6.2	234081	2	PFMAL4P2	AL035475 Plasmodiu
	26	61.2	6.1	110000	1	AP008934_03	Continuation (4 of
c	27	60.8	6.1	1522	2	AY701231	AY701231 Orconecte
	28	58.8	5.9	14255	1	AE010240	AE010240 Pyrococcu
c	29	58.4	5.8	176174	8	AC007483	AC007483 Homo sapi
c	30	57.2	5.7	168624	8	AC092131	AC092131 Homo sapi
c	31	57.2	5.7	173189	8	AC092323	AC092323 Homo sapi
	32	56.6	5.6	250743	2	AE014836	AE014836 Plasmodiu
	33	55	5.5	175946	8	AC141000	AC141000 Homo sapi
c	34	54.6	5.4	110000	14	PFMAL7P1_06	Continuation (7 of
	35	54.2	5.4	251448	2	AE014819	AE014819 Plasmodiu
	36	54	5.4	159910	8	AC009235	AC009235 Homo sapi
	37	53.8	5.4	253151	2	AE014842	AE014842 Plasmodiu
c	38	53.6	5.3	254733	2	AC117075	AC117075 Dictyoste
	39	53.4	5.3	110295	15	AC145164	AC145164 Medicago
	40	53.2	5.3	165165	14	CR628390	CR628390 Danio rer
	41	52.6	5.2	180861	14	CR391970	CR391970 Danio rer
	42	52.4	5.2	208848	14	AC139608	AC139608 Rattus no
c	43	52.4	5.2	283619	14	CR753823	CR753823 Danio rer
	44	52.2	5.2	189668	14	AC154966	AC154966 Bos tauru
	45	52.2	5.2	204551	14	AC158045	AC158045 Bos tauru

ALIGNMENTS

RESULT 1
SERFBB
LOCUS SERFBB 22080 bp DNA linear BCT 14-JUN-2005
DEFINITION S.enterica (group B) rfb gene cluster.
ACCESSION X56793 M29713
VERSION X56793.1 GI:47885
KEYWORDS abe gene; CDP-abequose synthase; CDP-glucose 4.6-dehydratase; ddhA gene; ddhB gene; ddhC gene; ddhD gene; dTDP-6-deoxy-L-mannose dehydrogenase; dTDP-dehydrorhamnose reductase; dTDP-glucose oxidoreductase; dTDP-glucose-4.6-dehydratase; galF gene; glucose-1-P cytidyltransferase; glucose-1-P thymidydyltransferase; manB gene; manC gene; Manose-1-Pguanylyltransferase; oxidoreductase; Phosphomanomutase; rfb gene cluster; rmlA gene; rmlB gene; rmlC gene; rmlD gene; undecaprenyl-P galactose-P transferase; wbaN gene; wbaP gene; wbaU gene; wbaV gene; wcaL gene; wzx gene.
SOURCE Salmonella enterica
ORGANISM Salmonella enterica
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 22080)
AUTHORS Wyk,P. and Reeves,P.
TITLE Identification and sequence of the gene for abequose synthase, which confers antigenic specificity on group B salmonellae: homology with galactose epimerase
JOURNAL J. Bacteriol. 171 (10), 5687-5693 (1989)
PUBMED 2793832
REFERENCE 2 (bases 1 to 22080)
AUTHORS Jiang,X.M., Neal,B., Santiago,F., Lee,S.J., Romana,L.K. and Reeves,P.R.
TITLE Structure and sequence of the rfb (O antigen) gene cluster of Salmonella serovar typhimurium (strain LT2)
JOURNAL Mol. Microbiol. 5 (3), 695-713 (1991)
PUBMED 1710759
REFERENCE 3 (bases 1 to 22080)
AUTHORS Reeves,P.
TITLE Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale
JOURNAL Trends Genet. 9 (1), 17-22 (1993)
PUBMED 8434412
REFERENCE 4 (bases 1 to 22080)

```
  AUTHORS   Liu,D., Haase,A.M., Lindqvist,L., Lindberg,A.A. and Reeves,P.R.
  TITLE     Glycosyl transferases of O-antigen biosynthesis in Salmonella
            enterica: identification and characterization of transferase genes
            of groups B, C2, and E1
  JOURNAL   J. Bacteriol. 175 (11), 3408-3413 (1993)
   PUBMED   7684736
REFERENCE   5  (bases 1 to 22080)
  AUTHORS   Stevenson,G., Andrianopoulos,K., Hobbs,M. and Reeves,P.R.
  TITLE     Organization of the Escherichia coli K-12 gene cluster responsible
            for production of the extracellular polysaccharide colanic acid
  JOURNAL   J. Bacteriol. 178 (16), 4885-4893 (1996)
   PUBMED   8759852
REFERENCE   6  (bases 1 to 22080)
  AUTHORS   Marolda,C.L. and Valvano,M.A.
  TITLE     The GalF protein of Escherichia coli is not a UDP-glucose
            pyrophosphorylase but interacts with the GalU protein possibly to
            regulate cellular levels of UDP-glucose
  JOURNAL   Mol. Microbiol. 22 (5), 827-840 (1996)
   PUBMED   8971705
REFERENCE   7  (bases 1 to 22080)
  AUTHORS   Reeves,P.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-1990) P.R. Reeves, G08 DEPT OF MICROBIOLOGY,
            UNIVERSITY OF SYDNEY, NSW 2006, AUSTRALIA
COMMENT     On Jun 13, 2005 this sequence version replaced gi:154337.
FEATURES             Location/Qualifiers
     source          1. .22080
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /strain="group B, serovar typhimurium, strain LT2"
                     /isolate="SL1654"
                     /db_xref="taxon:28901"
     gene            17. .1237
                     /gene="wcaL"
     CDS             17. .1237
                     /gene="wcaL"
                     /note="part of cluster for colanic acid prodaction"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA40132.1"
                     /db_xref="GI:47886"
                     /db_xref="GOA:P26388"
                     /db_xref="InterPro:IPR001296"
                     /db_xref="UniProt/Swiss-Prot:P26388"
                     /translation="MKVSPFLLKFPLSSETPVLNQITAFIDMGHEVEIVALQKGDTQH
                     THAAWEKYGLAAKTRWLQDEPQGRLAKLRYRACKTLPGLHRAATWKALNFTRYGDESR
                     NLILSAICAQVSQPFVADVFIAHFGPAGVTAAKLRELGVLRGKIATIFHGIDISSREV
                     LSHYTPEYQQLFRRGDLMLPISDLWAGRLKSMGCPPEKIAVSRMGVDMTRFTHRSVKA
                     PGMPLEMISVARLTEKKGLHVAIEACRQLKAQGVAFRYRILGIGPWERRLRTLIEQYQ
                     LEDVIEMPGFKPSHEVKAMLDDADVFLLPSITGTDGDMEGIPVALMEAMAVGIPVVST
                     VHSGIPELVEAGKSGWLVPENDAQALAARLAEFSRIDHDTLESVITRAREKVAQDFNQ
                     QAINRQLASLLQTI"
     gene            1248. .2651
                     /gene="ORF 1.3"
     CDS             1248. .2651
                     /gene="ORF 1.3"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA40113.1"
                     /db_xref="GI:47887"
                     /db_xref="GOA:P26389"
                     /db_xref="UniProt/Swiss-Prot:P26389"
                     /translation="MPATKFSRRTLLTAGSALAVLPFLRALPVQAREPRETVDIKDYP
                     ADDGIASFKQAFADGQTVVVPPGWVCENINAAITIPAGKTLRVQGAVRGNGRGRFILQ
                     DGCQVVGEQGGSLHNVTLDVRGSDCVIKGVAMSGFGPVAQIFIGGKEPQVMRNLIIDD
                     ITVTHANYAILRQGFHNQMDGARITHSRFSDLQGDAIEWNVAIHDRDILISDHVIERI
                     NCTNGKINWGIGIGLAGSTYDNSYPEDQAVKNFVVANITGSDCRQLVHVENGKHFVIR
                     NVKAKNITPGFSKNAGIDNATIAIYGCDNFVIDNIDMTNSAGMLIGYGVVKGKYLSIP
                     QNFKLNAIRLDNRQVAYKLRGIQISSGNTPSFVAITNVRMTRATLELHNQPQHLFLRN
                     INVMQTSAIGPALKMHFDLRKDVRGQFMARQDTLLSLANVHAINENGQSSVDIDRINH
                     QTVNVEAVNFSLPKRGG"
     gene            2829. .3722
                     /gene="galF"
     CDS             2829. .3722
                     /gene="galF"
                     /function="possibly in regulation of cellular levels of
                     UDP-glucose"
                     /codon_start=1
                     /transl_table=11
                     /product="putative glucose-1-phosphate
                     uridylyltransferase"
                     /protein_id="CAA40114.1"
                     /db_xref="GI:47888"
                     /db_xref="GOA:P0A2K7"
                     /db_xref="InterPro:IPR005774"
                     /db_xref="InterPro:IPR005835"
                     /db_xref="UniProt/Swiss-Prot:P0A2K7"
                     /translation="MMNLKAVIPVAGLGMHMLPATKAIPKEMLPIVDKPMIQYIVDEI
                     VAAGIKEIVLVTHASKNAVENHFDTSYELESLLEQRVKRQLLAEVQSICPPGVTIMNV
                     RQAQPLGLGHSILCARPVVGDNPFIVVLPDIIIDDATADPLRYNLAAMVARFNETGRS
                     QVLAKRMKGDLSEYSVIQTKEPLDNEGKVSRIVEFIEKPDQPQTLDSDLMAVGRYVLS
                     ADIWAELERTEPGAWGRIQLTDAIAELAKKQSVDAMLMTGDSYDCGKKMGYMQAFVKY
                     GLRNLKEGAKFRKSIEQLLHE"
     gene            4099. .5184
                     /gene="rmlB"
     CDS             4099. .5184
                     /gene="rmlB"
                     /EC_number="4.2.1.46"
                     /note="CDS starts with gtg"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="dTDP-glucose 4,6-dehydratase"
                     /protein_id="CAA40115.1"
                     /db_xref="GI:581654"
                     /db_xref="GOA:P26391"
                     /db_xref="InterPro:IPR001509"
                     /db_xref="InterPro:IPR005888"
                     /db_xref="UniProt/Swiss-Prot:P26391"
                     /translation="MKILITGGAGFIGSAVVRHIIKNTQDTVVNIDKLTYAGNLESLS
                     DISESNRYNFEHADICDSAEITRIFEQYQPDAVMHLAAESHVDRSITGPAAFIETNIV
                     GTYALLEVARKYWSALGEDKKNNFRFHHISTDEVYGDLPHPDEVENSVTLPLFTETTA
                     YAPSSPYSASKASSDHLVRAWRRTYGLPTIVTNCSNNYGPYHFPEKLIPLVILNALEG
                     KPLPIYGKGDQIRDWLYVEDHARALHMVVTEGKAGETYNIGGHNEKKNLDVVFTICDL
                     LDEIVPKATSYREQITYVADRPGHDRRYAIDAGKISRELGWKPLETFESGIRKTVEWY
                     LANTQWVNNVKSGAYQSWIEQNYEGRQ"
     gene            5184. .6083
                     /gene="rmlD"
     CDS             5184. .6083
                     /gene="rmlD"
                     /note="incorrectly designated rfbA in reference [2];
                     dTDP-glucose oxidoreductase"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="dTDP-6-deoxy-L-mannose dehydrogenase"
                     /protein_id="CAA40116.1"
                     /db_xref="GI:47890"
                     /db_xref="GOA:P26392"
                     /db_xref="InterPro:IPR005913"
                     /db_xref="UniProt/Swiss-Prot:P26392"
                     /translation="MNILLFGKTGQVGWELQRSLAPVGNLIALDVHSKEFCGDFSNPK
                     GVAETVRKLRPDVIVNAAAHTAVDKAESEPELAQLLNATSVEAIAKAANETGAWVVHY
                     STDYVFPGTGDIPWQETDATSPLNVYGKTKLAGEKALQDNCPKHLIFRTSWVYAGKGN
                     NFAKTMLRLAKERQTLSVINDQYGAPTGAELLADCTAHAIRVALNKPEVAGLYHLVAG
                     GTTTWHDYAALVFDEARKAGITLALTELNAVPTSAYPTPASRPGNSRLNTEKFQRNFD
                     LILPQWELGVKRMLTEMFTTTTI"
     gene            6131. .7009
                     /gene="rmlA"
     CDS             6131. .7009
                     /gene="rmlA"
                     /note="referred to as orf6.1 in reference [2]"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose-1-phosphate thymidylyltransferase"
                     /protein_id="CAA40117.1"
```

```
                     /db_xref="GI:47891"
                     /db_xref="GOA:P26393"
                     /db_xref="InterPro:IPR005835"
                     /db_xref="InterPro:IPR005907"
                     /db_xref="UniProt/Swiss-Prot:P26393"
                     /translation="MKTRKGIILAGGSGTRLYPVTMAVSKQLLPIYDKPMIYYPLSTL
                     MLAGIRDILIISTPQDTPRFQQLLGDGSQWGLNLQYKVQPSPDGLAQAFIIGEEFIGH
                     DDCALVLGDNIFYGHDLPKLMEAAVNKESGATVFAYHVNDPERYGVVEFDQKGTAVSL
                     EEKPLQPKSNYAVTGLYFYDNSVVEMAKNLKPSARGELEITDINRIYMEQGRLSVAMM
                     GRGYAWLDTGTHQSLIEASNFIATIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPL
                     SKNDYGKYLLKMVKGL"
     gene            7010. .7561
                     /gene="rmlC"
     CDS             7010. .7561
                     /gene="rmlC"

  Query Match             100.0%;  Score 1002;  DB 1;  Length 22080;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-158;
  Matches 1002;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14059 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 14118

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14119 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 14178

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14179 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 14238

Qy        181 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14239 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 14298

Qy        241 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14299 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 14358

Qy        301 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14359 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 14418

Qy        361 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14419 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 14478

Qy        421 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14479 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 14538

Qy        481 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14539 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 14598

Qy        541 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14599 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 14658

Qy        601 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14659 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 14718

Qy        661 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14719 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 14778

Qy        721 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14779 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 14838

Qy        781 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14839 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 14898

Qy        841 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14899 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 14958

Qy        901 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14959 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 15018

Qy        961 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
              |||||||||||||||||||||||||||||||||||||||||
Db      15019 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 15060


RESULT 2
BD085557
LOCUS       BD085557               22080 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Nucleic acid molecules specific for bacterial antigens and uses
            thereof.
ACCESSION   BD085557
VERSION     BD085557.1  GI:22631167
KEYWORDS    JP 2001522248-A/4.
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22080)
  AUTHORS   Reeves,P.R. and Wang,L.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
  JOURNAL   Patent: JP 2001522248-A 4 13-NOV-2001;
            THE UNIVERSITY OF SYDNEY
COMMENT     OS   Salmonella enterica
            PN   JP 2001522248-A/4
            PD   13-NOV-2001
            PF   01-MAY-1998 JP 1998547537
            PR   01-MAY-1997 AU      PO 6545,22-JUL-1997 AU      PO 8162 PI
            PETER RICHARD REEVES,LEI WANG
            PC   C12N9/10,C12N9/90,C12N9/92,C12N15/54,C12N15/61 CC   Nucleic
            acid molecules specific for bacterial antigens and  CC
            uses thereof
            FH   Key             Location/Qualifiers
            FT   source          1. .22080
            FT                   /organism='Salmonella enterica'.
FEATURES             Location/Qualifiers
     source          1. .22080
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:28901"
ORIGIN

  Query Match             100.0%;  Score 1002;  DB 6;  Length 22080;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-158;
  Matches 1002;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14059 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 14118

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14119 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 14178

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14179 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 14238

Qy        181 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14239 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 14298

Qy        241 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 300
```

```
Db      14299 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 14358

Qy        301 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 360

Db      14359 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 14418

Qy        361 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 420

Db      14419 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 14478

Qy        421 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 480

Db      14479 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 14538

Qy        481 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 540

Db      14539 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 14598

Qy        541 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 600

Db      14599 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 14658

Qy        601 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 660

Db      14659 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 14718

Qy        661 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 720

Db      14719 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 14778

Qy        721 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 780

Db      14779 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 14838

Qy        781 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 840

Db      14839 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 14898

Qy        841 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 900

Db      14899 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 14958

Qy        901 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 960

Db      14959 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 15018

Qy        961 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002

Db      15019 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 15060
```

```
RESULT 3
AX029264
LOCUS       AX029264               22080 bp    DNA     linear   PAT 16-SEP-2000
DEFINITION  Sequence 4 from Patent WO9850531.
ACCESSION   AX029264
VERSION     AX029264.1  GI:10190115
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1
  AUTHORS   Wang,L. and Reeves,P.R.
  TITLE     Nucleic acid molecules specific for bacterial antigens and uses
            thereof
  JOURNAL   Patent: WO 9850531-A 4 12-NOV-1998;
            UNIV SYDNEY (AU) ; WANG LEI (AU) ; REEVES PETER RICHARD (AU)
FEATURES             Location/Qualifiers
     source          1. .22080
                     /organism="Salmonella enterica"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:28901"
ORIGIN
```

```
  Query Match           100.0%;  Score 1002;  DB 6;  Length 22080;
  Best Local Similarity 100.0%;  Pred. No. 5.6e-158;
  Matches 1002;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60

Db      14059 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 14118

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120

Db      14119 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 14178

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180

Db      14179 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 14238

Qy        181 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 240

Db      14239 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 14298

Qy        241 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 300

Db      14299 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 14358

Qy        301 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 360

Db      14359 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 14418

Qy        361 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 420

Db      14419 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 14478

Qy        421 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 480

Db      14479 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 14538

Qy        481 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 540

Db      14539 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 14598

Qy        541 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 600

Db      14599 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 14658

Qy        601 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 660

Db      14659 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 14718

Qy        661 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 720

Db      14719 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 14778

Qy        721 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 780

Db      14779 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 14838

Qy        781 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 840

Db      14839 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 14898

Qy        841 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 900

Db      14899 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 14958

Qy        901 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 960

Db      14959 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 15018

Qy        961 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
```

```
Db      15019 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 15060

RESULT 4
AE008792/c
LOCUS       AE008792               22043 bp    DNA     linear   BCT 09-AUG-2005
DEFINITION  Salmonella typhimurium LT2, section 96 of 220 of the complete
            genome.
ACCESSION   AE008792 AE006468
VERSION     AE008792.1  GI:16420610
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 22043)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22043)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES             Location/Qualifiers
     source          1. .22043
                     /organism="Salmonella typhimurium LT2"
                     /mol_type="genomic DNA"
                     /strain="LT2; SGSC 1412; ATCC 700720"
                     /db_xref="ATCC:700720"
                     /db_xref="taxon:99287"
                     /note="LT2"
     gene            complement(71. .1064)
                     /gene="wzzB"
                     /note="synonym: STM2079"
     CDS             complement(71. .1054)
                     /gene="wzzB"
                     /note="chain length determinant protein (polysaccharide
                     antigen chainregulator).  (SW:WZZB_SALTY)"
                     /codon_start=1
                     /transl_table=11
                     /product="regulator of length of O-antigen component of
                     lipopolysaccharide chains"
                     /protein_id="AAL20983.1"
                     /db_xref="GI:16420611"
                     /translation="MTVDSNTSSGRGNDPEQIDLIELLLQLWRGKMTIIVAVIIAILL
                     AVGYLMIAKEKWTSTAIITQPDAAQVATYTNALNVLYGGNAPKISEVQANFISRFSSA
                     FSALSEVLDNQKEREKLTIEQSVKGQALPLSVSYVSTTAEGAQRRLAEYIQQVDEEVA
                     KELEVDLKDNITLQTKTLQESLETQEVVAQEQKDLRIKQIEEALRYADEAKITQPQIQ
                     QTQDVTQDTMFLLGSDALKSMIQNEATRPLVFSPAYYQTKQTLLDIKNLKVTADTVHV
                     YRYVMKPTLPVRRDSPKTAITLVLAVLLGGMIGAGIVLGRNALRSYKPKAL"
     RBS             complement(1059. .1064)
                     /gene="wzzB"
                     /note="putative RBS for wzzB; RegulonDB:STMS1H002228"
     gene            complement(1197. .2363)
                     /gene="udg"
                     /note="synonym: STM2080"
     CDS             complement(1197. .2363)
                     /gene="udg"
                     /EC_number="1.1.1.22"
                     /note="UDP-glucose 6-dehydrogenase.  (SW:UDG_SALTY)"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-glucose/GDP-mannose dehydrogenase"
                     /protein_id="AAL20984.1"
                     /db_xref="GI:16420612"
                     /translation="MKITISGTGYVGLSNGLLIAQHHDVVALDIVPSRVELLNDRISP
                     IVDKEIQQFLKEDNIRFRATLDKFDAYQNADYVIIATPTDYDPKTNYFNTSSVESVMQ
                     DVISINPAAVMIIKSTVPVGFTAAMRQKFATENIIFSPEFLREGKALYDNLYPSRIVI
                     GEQSERAREFAALLQEGAIKQEIPTLFTDSTEAEAIKLFANTYLAMRVAYFNELDSYA
                     ETLGLNTRQIIEGVCLDPRIGNHYNNPSFGYGGYCLPKDTKQLLANYQSVPNNIISAI
                     VEANRTRKDFIADAILARKPKVVGIYRLIMKSGSDNFRASSIQGIMKRIKAKGVEVII
                     YEPVMEEDTFFNSRLERDLHCFKQQADVIISNRMAAELLDVAEKVYTRDLFGSD"
     gene            complement(2600. .4018)
                     /gene="gnd"
                     /note="synonym: STM2081"
     CDS             complement(2600. .4006)
                     /gene="gnd"
                     /EC_number="1.1.1.44"
                     /note="decarboxylating; 6-phosphogluconate dehydrogenase,
                     decarboxylating.  (SW:6PGD_SALTY)"
                     /codon_start=1
                     /transl_table=11
                     /product="gluconate-6-phosphate dehydrogenase"
                     /protein_id="AAL20985.1"
                     /db_xref="GI:16420613"
                     /translation="MSKQQIGVVGMAVMGRNLALNIESRGYTVSVFNRSREKTEEVIA
                     ENPGKKLVPYYTVKEFVESLETPRRILLMVKAGAGTDAAIDSLKPYLEKGDIIIDGGN
                     TFFQDTIRRNRELSAEGFNFIGTGVSGGEEGALKGPSIMPGGQKDAYELVAPILTKIA
                     AVAEDGEPCVTYIGADGAGHYVKMVHNGIEYGDMQLIAEAYSLLKGGLNLSNEELANT
                     FTEWNNGELSSYLIDITKDIFTKKDEDGNYLVDVILDEAANKGTGKWTSQSALDLGEP
                     LSLITESVFARYISSLKAQRVAASKVLSGPKAQPAGDKAEFIEKVRRALYLGKIVSYA
                     QGFSQLRAASDEYHWDLNYGEIAKIFRAGCIIRAQFLQKITDAYAENADIANLLLAPY
                     FKKIADEYQQALRDVVAYAVQNGIPVPTFSAAVAYYDSYRAAVLPANLIQAQRDYFGA
                     HTYKRTDKEGIFHTEWLE"
     RBS             complement(4013. .4018)
                     /gene="gnd"
                     /note="putative RBS for gnd; RegulonDB:STMS1H002229"
     gene            complement(4170. .5607)
                     /gene="rfbP"
                     /note="synonym: STM2082"
     CDS             complement(4170. .5600)
                     /gene="rfbP"
                     /note="bifunctional enzyme; LPS side chain defect;
                     undecaprenyl-phosphate galactosephosphotransferase.
                     (SW:RFBP_SALTY); O-antigen transfer protein"
                     /codon_start=1
                     /transl_table=11
                     /product="undecaprenol-phosphate
                     galactosephosphotransferase, and"
                     /protein_id="AAL20986.1"
                     /db_xref="GI:16420614"
                     /translation="MDNIDNKYNPQLCKIFLAISDLIFFNLALWFSLGCVYFIFDQVQ
                     RFIPQDQLDTRVITHFILSVVCVGWFWIRLRHYTYRKPFWYELKEIFRTIVIFAIFDL
                     ALIAFTKWQFSRYVWVFCWTFALILVPFFRALTKHLLNKLGIWKKKTIILGSGQNARG
```

```
                AYSALQSEEMMGFDVIAFFDTDASDAEINMLPVIKDTEIIWDLNRTGDVHYILAYEYT
                ELEKTHFWLRELSKHHCRSVTVVPSFRGLPLYNTDMSFIFSHEVMLLRIQNNLAKRSS
                RFLKRTFDIVCSIMILLIIASPLMIYLWYKVTRDGGPAIYGHQRVGRHGKLFPCYKFRS
                MVMNSQEVLKELLANDPIARAEWEKDFKLKNDPRITAVGRFIRKTSLDELPQLFNVLK
                GDMSLVGPRPIVSDELERYCDDVDYYLMAKPGMTGLWQVSGRNDVDYDTRVYFDSWYV
                KNWTLWNDIAILFKTAKVVLRRDGAY"
     RBS        complement(5602. .5607)
                /gene="rfbP"
                /note="putative RBS for rfbP; RegulonDB:STMS1H002230"
     gene       complement(5672. .7112)
                /gene="rfbK"
                /note="synonym: STM2083"
     CDS        complement(5672. .7105)
                /gene="rfbK"
                /EC_number="5.4.2.8"
                /EC_number="1.1.1.44"
                /note="LPS side chain defect: phosphomannomutase.
                (SW:RFBK_SALTY)"
                /codon_start=1
                /transl_table=11
                /product="phosphomannomutase"
                /protein_id="AAL20987.1"
                /db_xref="GI:16420615"
                /translation="MNVVNNSRDVIYSSGIVFGTSGARGLVKDFTPQVCAAFTVSFVA
                VMQEHFSFDTVALAIDNRPSSYGMAQACAAALADKGVNCIFYGVVPTPALAFQSMSDN
                MPAIMVTGSHIPFERNGLKFYRPDGEITKHDEAAILSVEDTCSHLELKELIVSEMAAV
                NYISRYTSLFSTPFLKNKRIGIYEHSSAGRDLYKPLFIALGAEVVSLGRSDNFVPIDT
                EAVSKEDREKARSWAKEFDLDAIFSTDGDGDRPLIADEAGEWLRGDILGLLCSLALDA
                EAVAIPVSCNSIISSGRFFKHVKLTKIGSPYVIEAFNELSRSYSRIVGFEANGGFLLG
                SDICINEQNLHALPTRDAVLPAIMLLYKSRNTSISALVNELPTRYTHSDRLQGITTDK
                SQSLISMGRENLSNLLSYIGLENEGAISTDMTDGMRITLRDGCIVHLRASGNAPELRC
                YAEANLLNRAQDLVNTTLANIKKRCLL"
     gene       complement(7092. .8534)
                /gene="rfbM"
                /note="synonym: STM2084"
     CDS        complement(7092. .8531)
                /gene="rfbM"
                /EC_number="2.7.7.22"
                /EC_number="1.1.1.44"
                /note="LPS side chain defect: mannose-1-phosphate
                guanylyltransferase GDP).  (SW:RFBM_SALTY)"
                /codon_start=1
                /transl_table=11
                /product="mannose-1-phosphate guanylyltransferase"
                /protein_id="AAL20988.1"
                /db_xref="GI:16420616"
                /translation="MSFLPVIMAGGTGSRLWPLSREYHPKQFLSVEGKLSMLQNTIKR
                LASLSTEEPVVICNDRHRFLVAEQLREIDKLANNIILEPVGRNTAPAIALAAFCALQN
                ADNADPLLLVLAADHVIQDEIAFTKAVRHAEEYAANGKLVTFGIVPTHAETGYGYIRR
                GELIGNDAYAVAEFVEKPDIDTAGDYFKSGKYYWNSGMFLFRASSYLNELKYLSPEIY
                KACEKAVGHINPDLDFIRIDKEEFMSCPSDSIDYAVMEHTQHAVVIPMSAGWSDVGSW
                SSLWDISNKDHQRNVLKGDIFAHACNDNYIYSEDMFISAIGVSNLVIVQTTDALLVAN
                KDTVQDVKKIVDYLKRNDRNEYKQHQEVFRPWGKYNVIDSGKNYLVRCITVKPGEKFV
                AQMHHHRAEHWIVLSGTARVTKGEQTYMVSENESTFIPPNTIHALENPGMTPLKLIEI
                QSGTYLGEDDIIRLEQRSGFSKEWTNERS"
     RBS        complement(7107. .7112)
                /gene="rfbK"
                /note="putative RBS for rfbK; RegulonDB:STMS1H002231"
     RBS        complement(8529. .8534)
```

```
Query Match             99.8%;  Score 1000.4;  DB 1;  Length 22043;
Best Local Similarity   99.9%;  Pred. No. 1e-157;
Matches 1001;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      11859 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAAGAATATCTTGAAGAGTTGTTG 11800

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11799 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 11740

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11739 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 11680

Qy        181 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11679 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 11620

Qy        241 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11619 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 11560

Qy        301 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11559 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 11500

Qy        361 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11499 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 11440

Qy        421 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11439 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 11380

Qy        481 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11379 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 11320

Qy        541 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11319 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 11260

Qy        601 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11259 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 11200

Qy        661 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11199 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 11140

Qy        721 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11139 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 11080

Qy        781 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11079 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 11020

Qy        841 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11019 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 10960

Qy        901 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10959 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 10900

Qy        961 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
              |||||||||||||||||||||||||||||||||||||||||||
Db      10899 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 10858
```

```
RESULT 5
AY062284
LOCUS       AY062284               3651 bp    DNA     linear   BCT 20-MAR-2002
DEFINITION  Salmonella enterica subsp. enterica serovar Schleissheim O-antigen
            gene cluster, partial sequence.
ACCESSION   AY062284
VERSION     AY062284.1  GI:19550983
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Schleissheim
  ORGANISM  Salmonella enterica subsp. enterica serovar Schleissheim
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 3651)
```

```
  AUTHORS   Wang,L., Andrianopoulos,K., Liu,D., Popoff,M.Y. and Reeves,P.R.
  TITLE     Extensive variation in the O-antigen gene cluster within one
            Salmonella enterica serogroup reveals an unexpected complex history
  JOURNAL   J. Bacteriol. 184 (6), 1669-1677 (2002)
   PUBMED   11872718
REFERENCE   2  (bases 1 to 3651)
  AUTHORS   Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2001) Microbiology, The University of Sydney,
            Sydney, NSW 2006, Australia
FEATURES             Location/Qualifiers
     source          1. .3651
                     /organism="Salmonella enterica subsp. enterica serovar
                     Schleissheim"
                     /mol_type="genomic DNA"
                     /serovar="Schleissheim"
                     /sub_species="enterica"
                     /db_xref="taxon:181736"
     gene            <1. .983
                     /gene="wbaV"
     CDS             <1. .983
                     /gene="wbaV"
                     /note="abequosyl-transferase"
                     /codon_start=3
                     /transl_table=11
                     /product="WbaV"
                     /protein_id="AAL91078.1"
                     /db_xref="GI:19550984"
                     /translation="PTYNRKEYLEELLNSINNQENFNLDIEICISDNASTDGTEEMID
                     VWRNNYNFPIIYRRNSVNLGPDRNFLASVSLANGDYCWIFGSDDALAKDSLAILQTYL
                     DSQADIYLCDRKETGSDLAEIRSPHRSWLRTDDELYVFNNDLDREIYLSRCLSIGGVF
                     SYLSSLIVKKERWDAIDFDASYIGTSYPHVFIMMSVFNTPGCLLHYISKPLVICRGDN
                     DSFEKKGKARRILIDFIAYLKLANYFYSKNISLKRAFENVLLKERPWLYTTLAMACYG
                     NSDEKRDLSEFYAKLGCNKNMINTVLRFGKLAYAVKNITVLKNLTKRIIK"
     misc_feature    984. .>3651
                     /note="wbaV-wbaU intergenic region"
     gene            2398. .3573
                     /gene="wzy"
     CDS             2398. .3573
                     /gene="wzy"
                     /note="O-antigen polymerase"
                     /codon_start=1
                     /transl_table=11
                     /product="Wzy"
                     /protein_id="AAL91079.1"
                     /db_xref="GI:19550985"
                     /translation="MLPHTFMFLIFNIVYMLPVAYVHFFGFSLGGVEEIFNGDPSMLA
                     DMMVFYVAVLFFYLAGSALLGFGIKNKNIQWVIPFELPIIVKILFSLVLFSIVLSKIL
                     LYPEGVYSSYAFDSGAMASRVWNISMGLSELGIIIFTFCLVTKNIKFAVVTFLIVSLN
                     LLHGTRIFTLILLLMIIFYIIFINKRISKLKLFIILSLFFSAIVISFLIIFIIRSNVS
                     IDNINIDLILSPIVYESLFNQISFMRMLDWLHQGAVPFAPQMLFSDSAIFTLPTFLFD
                     EKSSLMYINKFGELSPLGGLSGYASAVIYFSYYYFLWYFILGITSSLLLRFTSSVNFV
                     ILSRVIYIYFVCDSLFRFNRDPWFIVIKMFVNNIVFILFVLMIIALKMRMEKRRAIQ"
ORIGIN

  Query Match             94.8%;  Score 949.4;  DB 1;  Length 3651;
  Best Local Similarity   97.9%;  Pred. No. 5.9e-149;
  Matches  962;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy         20 TTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATAGTATAAATAATCAGG 79
              |||||||||||||||||||| |||| ||||||||||||||||||||||||||||||||
Db          1 TTCCAACTTATAATAGAAAAGAATACCTTGAAGAGTTGTTGAATAGTATAAATAATCAGG 60

Qy         80 AAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCTCTACTGATGGTACAG 139
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAAATTTTAATTTAGATATTGAGATATGTATATCAGATAATGCCTCTACTGATGGTACAG 120

Qy        140 AGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCAATAATATATCGGCGTAATA 199
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        121 AGGAAATGATTGATGTTTGGAGGAATAATTATAATTTCCCAATAATATATCGGCGTAATA 180

Qy        200 GCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTATCCCTTGCGAATGGGGATT 259
              |||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||

Db        181 GCGTTAACCTTGGACCGGATAGGAATTTTCTTGCTTCAGTATCCCTTGCGAATGGGGATT 240

Qy        260 ATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGACTCGTTAGCGATATTACAAA 319
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        241 ATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGACTCTTTAGCGATATTACAAA 300

Qy        320 CTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGAAAAGAGACCGGGTGTGATT 379
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        301 CTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGAAAAGAGACCGGGAGTGATT 360

Qy        380 TAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACAGATGATGAACTTTATGTGT 439
              ||| |||||||||||  |||||||||| |||||||||||||||||||| ||||||||||
Db        361 TAGCTGAGATTAGAAGTCCTCATCGTTCCTGGCTCAGAACAGATGATGAGCTTTATGTGT 420

Qy        440 TTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATCTATTGGTGGTGTAT 499
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTAATAATGATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATCTATTGGTGGTGTAT 480

Qy        500 TTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGGGATGCCATTGATTTTGATG 559
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        481 TTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGCTGGGATGCCATTGATTTTGATG 540

Qy        560 CGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGCGTATTTAATACGC 619
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGCGTATTTAATACGC 600

Qy        620 CAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATATGCCGAGGAGATAATGATA 679
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATATGCCGAGGAGATAATGATA 660

Qy        680 GTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAAT 739
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        661 GTTTCGAGAAGAAAGGAAAGGCCAGACGAATATTAATTGATTTTATTGCATATTTAAAAT 720

Qy        740 TAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGC 799
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TAGCTAATTATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGC 780

Qy        800 TAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATG 859
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        781 TAAAAGAGAGACCATGGTTATATACAACTCTGGCTATGGCATGTTATGGCAATAGTGATG 840

Qy        860 AAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATATGATCAACA 919
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAAAAAGAGACTTATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATATGATCAACA 900

Qy        920 CTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATATTACCGTGCTTAAGAATT 979
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATATTACCGTGCTTAAGAATC 960

Qy        980 TTACTAAACGGATAATTAAGTAG 1002
              |||||||||||||||||||||||
Db        961 TTACTAAACGGATAATTAAGTAG 983


RESULT 6
CP000026_08
WPCOMMENT
  Sequence split into 46 fragments  LOCUS CP000026 Accession CP000026
     Fragment Name      Begin      End
     CP000026_00            1    110000
     CP000026_01       100001    210000
     CP000026_02       200001    310000
     CP000026_03       300001    410000
     CP000026_04       400001    510000
     CP000026_05       500001    610000
     CP000026_06       600001    710000
     CP000026_07       700001    810000
     CP000026_08       800001    910000
     CP000026_09       900001   1010000
     CP000026_10      1000001   1110000
     CP000026_11      1100001   1210000
     CP000026_12      1200001   1310000
```

```
    CP000026_13       1300001   1410000
    CP000026_14       1400001   1510000
    CP000026_15       1500001   1610000
    CP000026_16       1600001   1710000
    CP000026_17       1700001   1810000
    CP000026_18       1800001   1910000
    CP000026_19       1900001   2010000
    CP000026_20       2000001   2110000
    CP000026_21       2100001   2210000
    CP000026_22       2200001   2310000
    CP000026_23       2300001   2410000
    CP000026_24       2400001   2510000
    CP000026_25       2500001   2610000
    CP000026_26       2600001   2710000
    CP000026_27       2700001   2810000
    CP000026_28       2800001   2910000
    CP000026_29       2900001   3010000
    CP000026_30       3000001   3110000
    CP000026_31       3100001   3210000
    CP000026_32       3200001   3310000
    CP000026_33       3300001   3410000
    CP000026_34       3400001   3510000
    CP000026_35       3500001   3610000
    CP000026_36       3600001   3710000
    CP000026_37       3700001   3810000
    CP000026_38       3800001   3910000
    CP000026_39       3900001   4010000
    CP000026_40       4000001   4110000
    CP000026_41       4100001   4210000
    CP000026_42       4200001   4310000
    CP000026_43       4300001   4410000
    CP000026_44       4400001   4510000
    CP000026_45       4500001   4585229
Continuation (9 of 46) of CP000026 from base 800001 (CP000026 Salmonella enterica subsp.

  Query Match             36.0%;  Score 360.4;  DB 1;  Length 110000;
  Best Local Similarity   60.0%;  Pred. No. 3.3e-51;
  Matches  601;  Conservative    0;  Mismatches  401;  Indels    0;  Gaps    0;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||    |||||||||||||| ||||| ||||||  ||   | |||   |||||||  |   |
Db      70976 ATGAAGGTATCATTTTGTATCCCAACGTATAATCGAGTAAAATTCATTGAAGACCTTCTT 71035

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              | ||||| ||||||||                ||| | | ||  |||||||  |||||||||
Db      71036 GAAAGTATTAATAATCAATCTTCTCACTCCTTAATTGTAGAAGTATGTATTTCAGATAAT 71095

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              || || |||||||| || || |||   ||  || | |||||  | |  | ||||||  |||
Db      71096 GCATCAACTGATGGAACTGAAGAATCAATAAATATCTGGAGAGATAGATTTAATTTTCCA 71155

Qy        181 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 240
              ||  ||||||    |  ||| ||  ||  ||||||||||||||||| | |   || ||||
Db      71156 ATTTTATATCAAAGGCATAACGAAAATATTGGGCCAGATAGGAATTATTTATCTGCAGTT 71215

Qy        241 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 300
              | | || ||| |||||||||||||| || || |||||||||  |||| | ||| |
Db      71216 AATATGGGGACTGGAGATTATTGTTGGATTTTCGGAAGTGATGATATTCTTACAAAAAAT 71275

Qy        301 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 360
              ||  | ||  |  |  ||  | |  | | | |    |   || || ||||||||||||||
Db      71276 TCCCTTGCATTGATGGAAGATAAATTGGCTGCCGGAAGTGACATTTATTTATGTGACAGA 71335

Qy        361 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 420
               |||     |    |     |     | ||    || || |||||    ||| | |
Db      71336 CGAGAATTAGATATTTCAATGACGAAAATATCTAATCCACATCGGCGATGGTTAAATGGT 71395

Qy        421 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 480
              |    | | | | |   |||  ||| |    ||| |  |   ||  | |||| |||
Db      71396 GGCAGTAGATTATTTTCTTTTAGCAATGAAGCTGATTTGATAGAGTACTTTAGTAAATGT 71455

Qy        481 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 540
                ||  | || ||  | ||||||||  |||| ||| | ||||| ||||   |   ||||
```

```
Db      71456 AACTCAGTCGGCGGACTTTTTAGCTACTTAAGCTCTATTATAGTGAAAAGGAATAAATGG 71515

Qy        541 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 600
              || |||    ||||| | ||| || ||||| ||  | ||| | ||||||| |||  |
Db      71516 TCTGACATCTTATTTGACGAGTCATACATTGGTACAGCTTATGCGCATGTATATATTTTA 71575

Qy        601 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 660
               | ||  || |||| |         |       |||| || ||||||   ||||| ||
Db      71576 TTAAGAATAATTAACAATATGAATTCGACCCTGCAGTACATATCACTGCCACTGGTTGAT 71635

Qy        661 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 720
              ||| |||||| ||| ||||  |||||  |   || || |||||   | |||||   |||||||
Db      71636 TGCAGAGGAGGTAACGATACATTCGAAAGTAATGGGAAGGCTCGTCGAATAAAAATTGAT 71695

Qy        721 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 780
              || || |  |||||||| |     | |||||||| | ||| |||| |    || |
Db      71696 TTCATCGGTTATTTAAAACTGAGGGAGGATTTTTATAATAACAATACAAAAATATATATC 71755

Qy        781 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 840
              ||||||  |   ||| |  ||||||| | || ||||| |||||  |||| |||||||||
Db      71756 TCATTTGGACGGGTTCTTACAAAAGAGCGTCCTTGGTTTTATACTTCTTTAGCTATGGCA 71815

Qy        841 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 900
              ||||||||  ||  |   || | |   ||  |  |   |||||||   ||| | || | |
Db      71816 TGTTATGGTGATTCTACCGATAGAGCTGAACTTGCATCATTTTATAAGAAGTTGGGATAT 71875

Qy        901 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 960
              ||||     | |||     || || |  ||       || ||             |||||
Db      71876 CCGAAAATAGCCACCAATTTAATATTTTGTCTTAAAGGGCTGGCTAGCTACACTAAAAAA 71935

Qy        961 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
              || |   |    || |   | |  |||   ||| ||   |||
Db      71936 ATAAAATTAGCAAAAATGGTAATAAAAAAAATATTTTCATAG 71977
```

```
RESULT 7
SERFBD1B
LOCUS       SERFBD1B               2879 bp    DNA     linear   BCT 12-MAR-1997
DEFINITION  Salmonella enterica group D1 wzx gene, partial cds and wbaV gene,
            complete cds.
ACCESSION   M65054
VERSION     M65054.1  GI:154334
KEYWORDS    .
SOURCE      Salmonella enterica
  ORGANISM  Salmonella enterica
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 2879)
  AUTHORS   Liu,D., Verma,N.K., Romana,L.K. and Reeves,P.R.
  TITLE     Relationships among the rfb regions of Salmonella serovars A, B,
            and D
  JOURNAL   J. Bacteriol. 173 (15), 4814-4819 (1991)
   PUBMED   1856174
REFERENCE   2  (bases 1 to 2879)
  AUTHORS   Reeves,P.
  TITLE     Evolution of Salmonella O antigen variation by interspecific gene
            transfer on a large scale
  JOURNAL   Trends Genet. 9 (1), 17-22 (1993)
   PUBMED   8434412
REFERENCE   3  (bases 1 to 2879)
  AUTHORS   Reeves,P.R., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C.,
            Coplin,D., Kido,N., Klena,J., Maskell,D., Raetz,C.R. and Rick,P.D.
  TITLE     Bacterial polysaccharide synthesis and gene nomenclature
  JOURNAL   Trends Microbiol. 4 (12), 495-503 (1996)
   PUBMED   9004408
FEATURES             Location/Qualifiers
     source          1. .2879
                     /organism="Salmonella enterica"
                     /mol_type="genomic DNA"
                     /strain="Ty2, group D1, serovar typhi"
                     /db_xref="taxon:28901"
                     /note="polysaccharide genes have been named in accordance
                     with the Bacterial Polysaccharide Gene Nomenclature
```

```
                scheme"
     gene       1. .1299
                /gene="wzx"
     CDS        <1. .1299
                /gene="wzx"
                /function="putative O-antigen exporter"
                /note="previously called rfbX; referred to as orf12.8 in
                citation [2]; bases 1-28 overlaps with bases 2283-2310 of
                M29682"
                /citation=[3]
                /codon_start=1
                /transl_table=11
                /product="putative 12 transmembrane segment protein"
                /protein_id="AAB49394.1"
                /db_xref="GI:154335"
                /translation="VRKLRLVRIPRHLIIAASSWLSKIIIAGVQLVSVKFLLEILGEE
                SYAVFTLLTGLLVWFSIADIGIGSSLQNYISELKADRKSYDAYIKAAVHILFASLIIL
                SSTLFFLSDKLSSLYLTSFSDELKNNSGSYFFIASILFIFIGVGSVVYKILFAELLGW
                KANIINALSYLLGFLDVVAIHYLMPDSSITFALVALYAPVAILPIIYISFRYIYVLKA
                KVNFNTYKLLLSRSSGFLIFSSLSIIVLQTDYIVMSQKLSAADIIKYTVTMKIFGLMF
                FIYTAVLQALWPVCAELRVKMQWRKLHRIIFLNIIGGVFFIGLGTLFIYVLKDYIYSI
                IANGIDYNISGVVFVLLAVYFSIRVWCDTFAMLLQSMNQLKILWLIVPCQALIGGVTQ
                WYFAEHYGIVGILYGLILSFSLTVFWGLPVYYMYKSKRLA"
     gene       1301. .2302
                /gene="wbaV"
     CDS        1301. .2302
                /gene="wbaV"
                /note="orf14.1; possibly encodes abequosyl transferase"
                /codon_start=1
                /transl_table=11
                /protein_id="AAB49395.1"
                /db_xref="GI:154336"
                /translation="MKVSFCIPTYNRVKFIEDLLESINNQSSHSLIVEVCISDNASTD
                GTEESININWRDRFNFPILYQRHNENIGPDRNYLSAVNMGTGDYCWIFGSDDILTKNSL
                ALMEDKLAAGSDIYLCDRRELDISMTKISNPHRRWLNGGSRLFSFSNEADLIEYFSKC
                NSVGGLFSYLSSIIVKRNKWSDVLFDESYIGTAYAHVYILLRIINNMNSTLQYISLPL
                VDCRGDNDTFESNGKARRIKIDFIGYLKLREDFYNNNTKIYISFGRVLTKERPWLYTS
                LAMACYGDSTDRAELASFYKKLGYPKIATNLIFCLKGLASYTKKIKLAKMVIKKIFS"
ORIGIN
```

Query Match 35.8%; Score 358.8; DB 1; Length 2879;
Best Local Similarity 59.9%; Pred. No. 2e-50;
Matches 600; Conservative 0; Mismatches 402; Indels 0; Gaps 0;

```
Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
Db       1301 ATGAAGGTATCATTTTGTATCCCAACGTATAATCGAGTAAAATTCATTGAAGACCTTCTT 1360

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
Db       1361 GAAAGTATTAATAATCAATCTTCTCACTCCTTAATTGTAGAAGTATGTATTTCAGATAAT 1420

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
Db       1421 GCATCAACTGATGGAACTGAAGAATCAATAAATATCTGGAGAGATAGATTTAATTTTCCA 1480

Qy        181 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 240
Db       1481 ATTTTATATCAAAGGCATAACGAAAATATTGGGCCAGATAGGAATTATTTATCTGCAGTT 1540

Qy        241 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 300
Db       1541 AATATGGGGACTGGAGATTATTGTTGGATTTTCGGAAGTGATGATATTCTTACAAAAAT 1600

Qy        301 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 360
Db       1601 TCCCTTGCATTGATGGAAGATAAATTGGCTGCCGGAAGTGACATTTATTTATGTGACAGA 1660

Qy        361 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 420
Db       1661 CGAGAATTAGATATTTCAATGACGAAAATATCTAATCCACATCGGCGATGGTTAAATGGT 1720

Qy        421 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 480
Db       1721 GGCAGTAGATTATTTTCTTTTAGCAATGAAGCTGATTTGATAGAGTACTTTAGTAAATGT 1780

Qy        481 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 540
Db       1781 AACTCAGTCGGCGGACTTTTTAGCTACTTAAGCTCTATTATAGTGAAAAGGAATAAATGG 1840

Qy        541 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 600
Db       1841 TCTGACGTCTTATTTGACGAGTCATACATTGGTACAGCTTATGCGCATGTATATATTTTA 1900

Qy        601 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 660
Db       1901 TTAAGAATAATTAACAATATGAATTCGACCCTGCAGTACATATCACTGCCACTGGTTGAT 1960

Qy        661 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 720
Db       1961 TGCAGAGGAGATAACGATACATTCGAAAGTAATGGGAAGGCTCGTCGAATAAAAATTGAT 2020

Qy        721 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 780
Db       2021 TTCATCGGTTATTTAAAACTGAGGGAGGATTTTTATAATAACAATACAAAAATATATATC 2080

Qy        781 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 840
Db       2081 TCATTTGGACGGGTTCTTACAAAAGAGCGTCCTTGGCTTTATACTTCTTTAGCTATGGCA 2140

Qy        841 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 900
Db       2141 TGTTATGGTGATTCTACCGATAGAGCTGAACTTGCATCATTTTATAAGAAGTTGGGATAT 2200

Qy        901 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 960
Db       2201 CCGAAAATAGCCACCAATTTAATATTTTGTCTTAAAGGGCTGGCTAGCTACACTAAAAAA 2260

Qy        961 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
Db       2261 ATAAAATTAGCAAAAATGGTAATAAAAAAAATATTTTCATAG 2302
```

```
RESULT 8
AL627273/c
LOCUS       AL627273              250050 bp    DNA     linear   BCT 16-APR-2005
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 9/20.
ACCESSION   AL627273 AL513382
VERSION     AL627273.1  GI:16503247
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 250050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
   PUBMED   11677608
REFERENCE   2  (bases 1 to 250050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
```

```
          Details of S. typhi sequencing at the Sanger Centre are available
          on the World Wide Web.
          (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .250050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            complement(109. .1566)
                     /gene="STY2216"
     CDS             complement(109. .1566)
                     /gene="STY2216"
                     /note="Orthologue of E. coli YEEO_ECOLI; Fasta hit to
                     YEEO_ECOLI (547 aa), 81% identity in 488 aa overlap
                     Unknown function. Contains hydrophobic, possible
                     membrane-spanning regions"
                     /codon_start=1
                     /transl_table=11
                     /product="putative inner membrane protein"
                     /protein_id="CAD02374.1"
                     /db_xref="GI:16503248"
                     /db_xref="GOA:Q8Z5P2"
                     /db_xref="InterPro:IPR002528"
                     /db_xref="UniProt/TrEMBL:Q8Z5P2"
                     /translation="MWFQLSSLRSILNVSTALRQVVVRTPWYAKRKSYKVLFWREITP
                     LAIPIFLENTCVLLMGVLSTFLVSWLGKEAMAGVGLADSFNMVIMAFFAAIDLGTTVV
                     VAFSLGKRDRRRARAAARQSLVIMTLFAVVLAVVIHYFGSEIINIVAGEATPEVKGLA
                     LTYLELTVLSYPAAAIALIGSGALRGAGNTKIPLMINGGMNILNIVISSILIYGAFSW
                     QGLGFAGAGLGLTISRYIGAVAIIWVLMIGFNPALRIPLKSYLKPLNFGIIWEVMGIG
                     IPASIESVLFNGGKLLTQMFVAGMGTNVIAGNFIAFSVAALINLPGNALGSASTIITG
                     KRLGTGQIGQAERQLRHVFWMSTIVLTAIAWGTAPFAGLFASFYTQEQDVKEVVKVLL
                     WLNAAFMPIWAAAWVLPSGFKGARDVRFAMWVSMLGMWGCRVVAGYTLGIVLGMGVVG
                     VWLGMFLDWAVRGALFYWRLISGRWLWRYLRVKRE"
     misc_feature    complement(199. .1044)
                     /gene="STY2216"
                     /note="Pfam match to entry PF01554 UPF0013,
                     Uncharacterized membrane protein family UPF0013, score
                     273.80, E-value 2.2e-78"
     tRNA            1726. .1801
                     /product="tRNA-Asn"
                     /note="tRNA Asn anticodon GTT, Cove score 87.13"
     gene            complement(1806. .1976)
                     /gene="STY2217"
     CDS             complement(1806. .1976)
                     /gene="STY2217"
                     /note="Unknown function"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAD02375.1"
                     /db_xref="GI:16503249"
                     /db_xref="UniProt/TrEMBL:Q8Z5P1"
                     /translation="MSVMAGASFSQRTARPPIKPPKTMDAKKPGKPYQVNRVKISSYF
                     LNVSGLAGISRN"
     gene            complement(2027. .2956)
                     /gene="STY2218"
                     /note="synonym: erfK"
     CDS             complement(2027. .2956)
                     /gene="STY2218"
                     /note="Similar to Salmonella typhimurium protein erfK/srfK
                     precursor SW:ERFK_SALTY (P40680) (309 aa) fasta scores:
                     E(): 0, 99.0% id in 309 aa. Contains probable N-terminal
                     signal sequence
                     Fasta hit to YNHG_ECOLI (334 aa), 47% identity in 288 aa
                     overlap
                     Fasta hit to YCFS_ECOLI (320 aa), 48% identity in 307 aa
                     overlap
                     Fasta hit to YBIS_ECOLI (306 aa), 66% identity in 283 aa
                     overlap
                     Orthologue of E. coli erfK (ERFK_ECOLI); Fasta hit to
                     ERFK_ECOLI (310 aa), 79% identity in 305 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative exported protein"
                     /protein_id="CAD02376.1"
                     /db_xref="GI:16503250"
                     /db_xref="InterPro:IPR005490"
                     /db_xref="UniProt/TrEMBL:Q8Z5P0"
                     /translation="MRRITPFFPFFVLLVSHFSLAISYPLPPEGSRLVGQPVTIAVPQ
                     NNTQPLESFAARYGQGLSNMLEANPGVDVFLPQSGSTLVVPQQLILPDTVREGIVVNV
                     AEMRLYYYPAGTNTVEVLPIGIGQAGRETPRNWVTAVERKQDGPVWVPTANTRREYAK
                     EGKTLPAMVPAGPDNPMGLYAIYIGRLYATHGTNANFGIGLRVSQGCIRLRNDDIKYL
                     FDHVRVGTRVQIIDRPVKFSVEPDGSRWLEVHEPLSRNRAEFESDKKVPLPVTPVLRT
                     FIKGDDVDTSRVNEVLERRSGMPVNISAGRSGL"
     RBS             complement(2964. .2970)
                     /note="possible RBS"
     gene            complement(3034. .4104)
                     /gene="STY2219"
                     /note="synonym: cobT"
     CDS             complement(3034. .4104)
                     /gene="STY2219"
                     /EC_number="2.4.2.21"
                     /note="Similar to Salmonella typhimurium
                     nicotinate-nucleotide--dimethylbenzimidazole
                     phosphoribosyltransferase SW:COBT_SALTY (Q05603) (356 aa)
                     fasta scores: E(): 0, 97.8% id in 356 aa
                     Orthologue of E. coli cobT (COBT_ECOLI); Fasta hit to
                     COBT_ECOLI (359 aa), 78% identity in 356 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="nicotinate-nucleotide--dimethylbenzimidazole
                     phosphoribosyl transferase"
                     /protein_id="CAD02377.1"
                     /db_xref="GI:16503251"
                     /db_xref="GOA:Q8Z5N9"
                     /db_xref="InterPro:IPR003200"
                     /db_xref="InterPro:IPR008281"
                     /db_xref="UniProt/Swiss-Prot:Q8Z5N9"
                     /translation="MQTLHALFRDIPAPDAEAMARAQQHIDGLLKPPGSLGRLETLAV
                     QLAGMPGLNGTPQVGEKAVLVMCADHGVWDEGVAVSPKIVTAIQAANMTQGTTGVCVL
                     AAQAGAKVHVIDVGIDAEPIPGVVDMRVARGCGNIAVGPAMSRSQAEALLLEVSRYTC
                     DLAKRGVTLFGVGELGMANTTPAAAMVSVFTGSDAKEVVGIGANLPPSRIDNKVDVVR
                     RAIAINQPNPRDGIDVLSKVGGPDLVGMTGVMLGAARCGLPVLLDGFLSYSAALAACQ
                     IAPAVRPYLIPSHFSAEKGARIALAHLSMEPYLHMAMRLGEGSGAALAMPIVEAACAM
                     FHNMGELAASNIVLPEGNANAT"
     misc_feature    complement(3067. .4065)
                     /gene="STY2219"
                     /note="Pfam match to entry PF02277 DBI_PRT,
                     Phosphoribosyltransferase, score 794.90, E-value 3.1e-235"
     gene            complement(4131. .4874)
                     /gene="STY2220"
                     /note="synonym: cobS"
     CDS             complement(4131. .4874)
                     /gene="STY2220"
                     /EC_number="2.-.-.-"
                     /note="Similar to Salmonella typhimurium cobalamin
                     [5'-phosphate] synthase cobS SW:COBS_SALTY (Q05602) (247
                     aa) fasta scores: E(): 0, 97.6% id in 247 aa
                     Orthologue of E. coli cobS (COBS_ECOLI); Fasta hit to
                     COBS_ECOLI (247 aa), 82% identity in 247 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="cobalamin (5'-phosphate) synthase"
                     /protein_id="CAD02378.1"
                     /db_xref="GI:16503252"
                     /db_xref="GOA:Q8Z5N8"
                     /db_xref="InterPro:IPR003805"
                     /db_xref="UniProt/Swiss-Prot:Q8Z5N8"
                     /translation="MSKLFWAMLAFISRLPVPSRWSQGLDFEQYSRGIVMFPFIGLIL
                     GGISGLIFILLQPWCGIPLAALFCILALALLTGGFHLDGLADTCDGIFSARRRERMLE
                     IMRDSRLGTHGGLALIFVLLAKILVVSELALRGTPMLAALAAACAAGRGSAVLLMYRH
                     RYAREEGLGNVFIGKVSGRQTCITLGLAVIVATVLLLGMQGLATMVVTLAAIFILGQL
                     LKRTLGGQTGDTLGAAIELGELIFLLALL"
     gene            complement(4871. .5416)
```

```
                     /gene="STY2221"
                     /note="synonym: cobU"
     CDS             complement(4871. .5416)
                     /gene="STY2221"
                     /note="Similar to Salmonella typhimurium bifunctional
                     cobalamin biosynthesis protein cobP [includes: cobinamide
                     kinase; cobinamide phosphate guanylyltransferase] cobU
                     SW:COBU_SALTY (Q05599) (181 aa) fasta scores: E(): 0,
                     99.4% id in 181 aa
                     Orthologue of E. coli cobU (COBU_ECOLI); Fasta hit to
                     COBU_ECOLI (181 aa), 82% identity in 181 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="cobinamide kinase and guanylyltransferase"
                     /protein_id="CAD02379.1"
                     /db_xref="GI:16503253"
                     /db_xref="GOA:Q8Z5N7"
                     /db_xref="InterPro:IPR003203"
                     /db_xref="UniProt/TrEMBL:Q8Z5N7"
                     /translation="MMILVTGGARSGKSRHAEALISDAPQVLYIATSQILDDEMAARI
                     QHHKDGRPAHWRTAECWRHLDTLITADLAPDDAILLECITTMVTNLLFALGGENDPEQ
                     WDYAAMERAIDDEIQILIAACQRCPAKVVLVTNEVGMGIVPENRLARHFRDIAGRVNQ
                     RLAAAADEVWLVVSGIGVKIK"

  Query Match            35.8%;  Score 358.8;  DB 1;  Length 250050;
  Best Local Similarity  59.9%;  Pred. No. 4.6e-51;
  Matches  600;  Conservative    0;  Mismatches  402;  Indels    0;  Gaps    0;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||    |||||||||||||| ||||| |||||| ||  | |||   |||||||  |  |
Db      74656 ATGAAGGTATCATTTTGTATCCCAACGTATAATCGAGTAAAATTCATTGAAGACCTTCTT 74597

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
               | ||||| ||||||||           |||  | | ||  |||||| |||||||||
Db      74596 GAAAGTATTAATAATCAATCTTCTCACTCCTTAATTGTAGAAGTATGTATTTCAGATAAT 74537

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              || || |||||||| || || |||    ||  || | |||||  | |  | ||||||| |||
Db      74536 GCATCAACTGATGGAACTGAAGAATCAATAAATATCTGGAGAGATAGATTTAATTTTCCA 74477

Qy        181 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 240
              ||  ||||||   |  ||| ||  ||  |||||||||||||||||| | |  || ||||
Db      74476 ATTTTATATCAAAGGCATAACGAAAATATTGGGCCAGATAGGAATTATTTATCTGCAGTT 74417

Qy        241 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 300
                | | || ||| |||||||||||||| || || |||||||||  |||| | ||| |
Db      74416 AATATGGGGACTGGAGATTATTGTTGGATTTTCGGAAGTGATGATATTCTTACAAAAAAT 74357

Qy        301 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 360
              ||  | ||  |  |  ||  | |  | | | |    |   || || ||||||||||||||
Db      74356 TCCCTTGCATTGATGGAAGATAAATTGGCTGCCGGAAGTGACATTTATTTATGTGACAGA 74297

Qy        361 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 420
              |||    |    |    |      | ||    || || |||||    ||| | |
Db      74296 CGAGAATTAGATATTTCAATGACGAAAATATCTAATCCACATCGGCGATGGTTAAATGGT 74237

Qy        421 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 480
              |    | | | | |   ||||  ||| |   |||  |  |   ||  | |||| |||
Db      74236 GGCAGTAGATTATTTTCTTTTAGCAATGAAGCTGATTTGATAGAGTACTTTAGTAAATGT 74177

Qy        481 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 540
              ||  | || ||  | |||||||  |||| ||| | ||||| ||||    |    ||||
Db      74176 AACTCAGTCGGCGGACTTTTTAGCTACTTAAGCTCTATTATAGTGAAAAGGAATAAATGG 74117

Qy        541 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 600
              || | |    ||||| | ||| || ||||| ||  | ||| | ||||||| |||  |
Db      74116 TCTGACGTCTTATTTGACGAGTCATACATTGGTACAGCTTATGCGCATGTATATATTTTA 74057

Qy        601 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 660
              | ||  || |||| |       |      |||| || ||||||   ||||| ||
Db      74056 TTAAGAATAATTAACAATATGAATTCGACCCTGCAGTACATATCACTGCCACTGGTTGAT 73997

Qy        661 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 720
              ||| |||||||||| ||||  ||||| |  || || |||||  | |||||   |||||||
Db      73996 TGCAGAGGAGATAACGATACATTCGAAAGTAATGGGAAGGCTCGTCGAATAAAAATTGAT 73937

Qy        721 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 780
              || || |  |||||||| |      | ||||||| | ||| |||| |    || |
Db      73936 TTCATCGGTTATTTAAAACTGAGGGAGGATTTTTATAATAACAATACAAAAATATATATC 73877

Qy        781 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 840
              |||||| |   ||| |   ||||||| | || ||| | |||||  |||| |||||||||
Db      73876 TCATTTGGACGGGTTCTTACAAAAGAGCGTCCTTGGCTTTATACTTCTTTAGCTATGGCA 73817

Qy        841 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 900
              ||||||||  ||  |   || | |   ||  |  |   |||||||   ||| | || | |
Db      73816 TGTTATGGTGATTCTACCGATAGAGCTGAACTTGCATCATTTTATAAGAAGTTGGGATAT 73757

Qy        901 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 960
                ||||    | |||    || || |  ||      || ||          |||||
Db      73756 CCGAAAATAGCCACCAATTTAATATTTTGTCTTAAAGGGCTGGCTAGCTACACTAAAAAA 73697

Qy        961 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
              || |   |   || |   | |  |||   ||| ||   |||
Db      73696 ATAAAATTAGCAAAAATGGTAATAAAAAAAATATTTTCATAG 73655


RESULT 9
AE016836
LOCUS       AE016836               300928 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 3 of
            16 of the complete genome.
ACCESSION   AE016836 AE014613
VERSION     AE016836.1  GI:29136667
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 300928)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
   PUBMED   12644504
REFERENCE   2  (bases 1 to 300928)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .300928
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(69. .1271)
                     /gene="ackA"
                     /locus_tag="t0527"
     CDS             complement(69. .1271)
                     /gene="ackA"
                     /locus_tag="t0527"
                     /note="corresponds to STY2567 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="acetate kinase"
                     /protein_id="AAO68233.1"
                     /db_xref="GI:29136668"
                     /translation="MSSKLVLVLNCGSSSLKFAIIDAVNGDEYLSGLAECFHLPEARI
                     KWKMDGSKQEAALGAGAAHSEALNFIVNTILAQKPELSAQLTAIGHRIVHGGEKYTSS
```

```
                VVIDESVIQGIKDSASFAPLHNPAHLIGIAEALKSFPQLKDKNVAVFDTAFHQTMPEE
                SYLYALPYSLYKEHGVRRYGAHGTSHFYVTQEAAKMLNKPVEELNIITCHLGNGGSVS
                AIRNGKCVDTSMGLTPLEGLVMGTRSGDIDPAIIFHLHDTLGMSVDQINKMLTKESGL
                LGLTEVTSDCRYVEDNYATKEDAKRAMDVYCHRLAKYIGSYTALMDGRLDAVVFTGGI
                GENAAMVRELSLGKLGVLGFEVDHERNLAARFGKSGFINKEGTRPAVVIPTNEELVIA
                QDASRLTA"
gene            1611. .2066
                /locus_tag="t0528"
CDS             1611. .2066
                /locus_tag="t0528"
                /note="corresponds to STY2566 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAO68234.1"
                /db_xref="GI:29136669"
                /translation="MSTPDNRSVNFFSLFRRGQHYAKTWPMEKRLAPVFVENRVIRMT
                RYAIRFMPPVAVFTLCWQIALGGQLGPAVATALFALSLPMQGLWWLGKRSLTPLPPSI
                LNWFYEVRGKLQEAGQALAPVEGKPDYQALADTLKRAFKQLDKTFLDDL"
gene            2149. .2643
                /locus_tag="t0529"
CDS             2149. .2643
                /locus_tag="t0529"
                /note="corresponds to STY2565 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAO68235.1"
                /db_xref="GI:29136670"
                /translation="MEMTNAQRLILSNQYKMMTMLDPTNAERYRRLQTIIERGYGLQM
                RELDREFGELTEETCRTIIDIMEMYHALHVSWTNLKDTQAIDERRVTFLGFDAATEAR
                YLGYVRFMVNIEGRYTHFDAGTHGFNAQTPMWEKYQRMLNVWHACPRQYHLSANEINQ
                IINA"
gene            2636. .3313
                /locus_tag="t0530"
CDS             2636. .3313
                /locus_tag="t0530"
                /note="corresponds to STY2564 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative phosphatase"
                /protein_id="AAO68236.1"
                /db_xref="GI:29136671"
                /translation="MPEEAHVQCKGFLFDLDGTLVDSLPAVERAWCSWADRFNLAHDE
                VLGFIHGKQAITSLRHFMAGKSEAEIAAEFTRLEQIEATETAGITALPGAVDLLNHLN
                KAGIPWAIVTSGSMPVARARHQVAGLPAPEVFVTAERVKRGKPEPDAYLLGAQLLGLA
                PQECVVVEDAPAGVLSGLAAGCHVIAVNAPADTPRLADVDFALDSLTQLSVAKQPNGD
                VVVLRKT"
gene            3389. .5215
                /locus_tag="t0531"
CDS             3389. .5215
                /locus_tag="t0531"
                /note="corresponds to STY2563 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative sodium/sulphate transporter"
                /protein_id="AAO68237.1"
                /db_xref="GI:29136672"
                /translation="MNGELIWVLSLLAIAVVLFATGKVRMDAVALFVIVAFVLSGTLT
                LSEAFSGFSDPNVILIAALFIIGDGLVRTGVATVVGTWLVKMAGSSEIKMLVLLMITV
                AGLGAFMSSTGVVAIFIPVVLSVSMHMQTSPSRLMMPLSFAGLISGMMTLVATPPNLV
                VNSELLREGLHGFNFFSVTPLGVVVLALGIVYMLVMRFMLKGDAPGQQAGKRRTFRDL
                IREYRLTGRARRLAIRPGSPMVGQRLDDLKLRERYGANVIGVERWRRFRRVIVNVNGV
                SEFRARDVLLIDMSAAEVDLREFCAEQLLEPMVLRGEYFSDQALDVGMAEISLIPKSE
                LIGKSVREIAFRTRYGLNVVGLKRDGVALEGSLADEPLLMGDIILVVGNWKLISLLGQ
                KGRDFVVLNMPVEVSEASPAHSQAPHAIFCLVLMVALMLTDEIPNPSAAIIACLLMGK
                FRCIDAESAYKAIHWPSIILIVGMMPFALALQKTGGVSLVVQGLMDIGGGYGPYMMLG
                CLFVLCAVIGLFISNTATAVLMAPIALAAAKSMGVSPYPFAMAVAMAASAAFMTPVSS
                PVNTLVLGPGNYSFSDFVKLGVPFTLIVMAVCIVMIPMLFPF"
```

```
gene            complement(5307. .5906)
                /locus_tag="t0532"
CDS             complement(5307. .5906)
                /locus_tag="t0532"
                /note="corresponds to STY2562 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative oxetanocin A biosynthetic enzyme"
                /protein_id="AAO68238.1"
                /db_xref="GI:29136673"
                /translation="MKQSHFFAHLSRMKLINRWPLMRNVRTENVSEHSLQVAMVAHAL
                AAIKNRKFGGQLNAERIALLAMYHDASEVLTGDLPTPVKYFNSQIAQEYKAIEKIAQQ
                KLVDMAPDELHDIFAPLIDENAWSKEEQAIVKQADALCAYLKCLEELSAGNNEFRLAK
                TRLEKTLELRRSQEMDYFMAVFVPSFHLSLDEISQDSPL"
gene            complement(6000. .7214)
                /locus_tag="t0533"
CDS             complement(6000. .7214)
                /locus_tag="t0533"
                /note="corresponds to STY2561 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative aminotransferase"
                /protein_id="AAO68239.1"
                /db_xref="GI:29136674"
                /translation="MSPIEKSSKLENVCYDIRGPVLKEAKRLEEEGNKVLKLNIGNPA
                PFGFEAPDEILVDVIRNLPTAQGYCDSKGLYSARKAIMQHYQARGMRDVTVEDIYIGN
                GVSELIVQAMQALLNSGDEMLVPAPDYPLWTAAVSLSSGKAVHYLCDESSDWFPDLDD
                IRAKITPRTRGIVIINPNNPTGAVYSKELLMEIVNIAREHNLIIFADEIYDKILYDDA
                EHHSIAALAPDLLTITFNGLSKTYRVAGFRQGWMVLNGPKKHAKGYIEGLEMLASMRL
                CANVPAQHAIQTALGGYQSISEFILPGGRLYEQRNRAWELINDIPGVSCVKPRGALYM
                FPRIDAKRFNIHDDQKMVLDFLLQEKVLLVQGTAFNWPWPDHFRIVTLPREDDLEMAI
                NRFGRFLSGYHQ"
gene            8146. .9084
                /gene="lrhA"
                /locus_tag="t0534"
CDS             8146. .9084
                /gene="lrhA"
                /locus_tag="t0534"
                /note="corresponds to STY2560 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="NADH dehydrogenase operon transcriptional
                regulator"
                /protein_id="AAO68240.1"
                /db_xref="GI:29136675"
                /translation="MINANRPIINLDLDLLRTFVAVADLNTFAAAAAAVCRTQSAVSQ
                QMQRLEQLVGKELFARHGRNKLLSEHGIQLLGYARKILRFNDEACSSLMFSNLQGVLT
                IGASDESADTILPFLLSRISSVYPKLALDVRVKRNAYMVDMVKSQEVDLVVTTNQPHS
                LDCLNLRTSPTHWYCAAEYVLQRGEPVPLVLLDDPSPFRDMVLETLNAAGIPWRLAYV
                ASTLLAVRAAVKAGLGVTARPVEMMSPDLRVLGAADGLPPLLDTEYMLCRDPNSQNEL
                AMVIFHAMESYQNPWHYNQFSAEGGDDPLMVEGGFE"
gene            9164. .9316
                /locus_tag="t0535"
CDS             9164. .9316
                /locus_tag="t0535"
                /note="corresponds to STY2559 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAO68241.1"
                /db_xref="GI:29136676"
                /translation="MPICTRLTAQIVKVAVNQPVAIRSSMWAKNNQNIHFMNLHVKKL
                EKLLLF"
gene            9709. .10152
                /gene="nuoA"
                /locus_tag="t0536"
CDS             9709. .10152
                /gene="nuoA"
                /locus_tag="t0536"
```

/note="corresponds to STY2558 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase I chain A"
/protein_id="AAO68242.1"
/db_xref="GI:29136677"

Query Match 35.8%; Score 358.8; DB 1; Length 300928;
Best Local Similarity 59.9%; Pred. No. 4.4e-51;
Matches 600; Conservative 0; Mismatches 402; Indels 0; Gaps 0;

```
Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||    |||||||||||||| ||||| |||||| ||  | |||   |||||||  |  |
Db     277792 ATGAAGGTATCATTTTGTATCCCAACGTATAATCGAGTAAAATTCATTGAAGACCTTCTT 277851

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
               | ||||| ||||||||              |||  | | ||  |||||||| |||||||||
Db     277852 GAAAGTATTAATAATCAATCTTCTCACTCCTTAATTGTAGAAGTATGTATTTCAGATAAT 277911

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              || || |||||||| || || |||   ||  || | |||||  | |  | |||||| |||
Db     277912 GCATCAACTGATGGAACTGAAGAATCAATAAATATCTGGAGAGATAGATTTAATTTTCCA 277971

Qy        181 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 240
              ||  ||||||   |  ||| ||  ||  ||||||||||||||||| | |  || ||||
Db     277972 ATTTTATATCAAAGGCATAACGAAAATATTGGGCCAGATAGGAATTATTTATCTGCAGTT 278031

Qy        241 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 300
                | | || ||| |||||||||||||| || || |||||||||  |||| | ||| |
Db     278032 AATATGGGGACTGGAGATTATTGTTGGATTTTCGGAAGTGATGATATTCTTACAAAAAAT 278091

Qy        301 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 360
              ||  | ||  |  |  ||  | |  | | | |   |   || || ||||||||||||||
Db     278092 TCCCTTGCATTGATGGAAGATAAATTGGCTGCCGGAAGTGACATTTATTTATGTGACAGA 278151

Qy        361 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 420
              |||    |    |    |     | ||    || || |||||    ||| | |
Db     278152 CGAGAATTAGATATTTCAATGACGAAAATATCTAATCCACATCGGCGATGGTTAAATGGT 278211

Qy        421 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 480
              |    |  | | | |   ||||  ||| |    ||| |  |   ||  | |||| |||
Db     278212 GGCAGTAGATTATTTTCTTTTAGCAATGAAGCTGATTTGATAGAGTACTTTAGTAAATGT 278271

Qy        481 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 540
              ||  | || ||  | ||||||||  |||| ||| | ||||| ||||   |   ||||
Db     278272 AACTCAGTCGGCGGACTTTTTAGCTACTTAAGCTCTATTATAGTGAAAAGGAATAAATGG 278331

Qy        541 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 600
              || | |    ||||| | ||| || |||||  ||  | ||| | ||||||| |||  |
Db     278332 TCTGACGTCTTATTTGACGAGTCATACATTGGTACAGCTTATGCGCATGTATATATTTTA 278391

Qy        601 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 660
              | ||  || |||| |       |      |||| || ||||||   ||||| ||
Db     278392 TTAAGAATAATTAACAATATGAATTCGACCCTGCAGTACATATCACTGCCACTGGTTGAT 278451

Qy        661 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 720
              ||| |||||||||| |||| ||||| |  || || |||||  | |||||   |||||||
Db     278452 TGCAGAGGAGATAACGATACATTCGAAAGTAATGGGAAGGCTCGTCGAATAAAAATTGAT 278511

Qy        721 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 780
              || || |  |||||||| |      | |||||||| | ||| |||| |    || |
Db     278512 TTCATCGGTTATTTAAAACTGAGGGAGGATTTTTATAATAACAATACAAAAATATATATC 278571

Qy        781 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 840
              ||||||  |   ||| |   ||||||| | || ||| | |||||  |||| |||||||||
Db     278572 TCATTTGGACGGGTTCTTACAAAAGAGCGTCCTTGGCTTTATACTTCTTTAGCTATGGCA 278631

Qy        841 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 900
              ||||||||  ||  |   || | |   ||  |  |  |||||||   ||| | || | |
Db     278632 TGTTATGGTGATTCTACCGATAGAGCTGAACTTGCATCATTTTATAAGAAGTTGGGATAT 278691

Qy        901 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 960
                  |||||     | |||     || || |  ||      || ||          |||||
Db     278692 CCGAAAATAGCCACCAATTTAATATTTTGTCTTAAAGGGCTGGCTAGCTACACTAAAAAA 278751

Qy        961 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
              || |   |    || |   | |  |||   ||| ||   |||
Db     278752 ATAAAATTAGCAAAAATGGTAATAAAAAAAATATTTTCATAG 278793
```

RESULT 10
AY062287
LOCUS AY062287 4115 bp DNA linear BCT 20-MAR-2002
DEFINITION Salmonella enterica subsp. enterica serovar Sloterdijk O-antigen gene cluster, partial sequence.
ACCESSION AY062287
VERSION AY062287.1 GI:19550990
KEYWORDS .
SOURCE Salmonella enterica subsp. enterica serovar Sloterdijk
ORGANISM Salmonella enterica subsp. enterica serovar Sloterdijk
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 4115)
AUTHORS Wang,L., Andrianopoulos,K., Liu,D., Popoff,M.Y. and Reeves,P.R.
TITLE Extensive variation in the O-antigen gene cluster within one Salmonella enterica serogroup reveals an unexpected complex history
JOURNAL J. Bacteriol. 184 (6), 1669-1677 (2002)
PUBMED 11872718
REFERENCE 2 (bases 1 to 4115)
AUTHORS Wang,L.
TITLE Direct Submission
JOURNAL Submitted (14-NOV-2001) Microbiology, The University of Sydney, Sydney, NSW 2006, Australia
FEATURES Location/Qualifiers
source 1. .4115
/organism="Salmonella enterica subsp. enterica serovar Sloterdijk"
/mol_type="genomic DNA"
/serovar="Sloterdijk"
/sub_species="enterica"
/db_xref="taxon:181798"
misc_feature <1. .>4115
/note="wbaV-wbaU intergenic region"
gene 2914. .4089
/gene="wzy"
CDS 2914. .4089
/gene="wzy"
/note="O-antigen polymerase"
/codon_start=1
/transl_table=11
/product="Wzy"
/protein_id="AAL91082.1"
/db_xref="GI:19550991"
/translation="MLPHTFMFLIFNIVYMLPVAYVHFFGFSLGGVEEIFNGDPSMLA
DMMVFYVAVLFFYLAGSALLGFGIKNKNIQWVIAFELPVIVKILFSLVLFSIVLSKVL
LYPEGVYSSYAFDSGAMASRVWNISMGLSELGIIIFTFCLVTKNIKFAVVTFLIVSLN
LLHGTRIFTLILLLMIIFYIIFINKKISKLKLFIILSLFFSAIVISFLIIFIIRSNVS
IDNINIDLILSPIVYESLFNQISFMRMLDWLHQGAVPFAPQMLFSDSAIFTLPTFLFD
EKSSLMYINKFGELSPLGGLSGYASAVIYFSYYYFLWYFILGITSSLLLRFTSSVNFV
ILSRVIYIYFVCDSLFRFNRDPWFIVIKMFVNNIVFILFVLMIIALKMRMEKRRAIQ"
ORIGIN

Query Match 28.0%; Score 280.2; DB 1; Length 4115;
Best Local Similarity 98.9%; Pred. No. 2.2e-37;
Matches 282; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy        718 GATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAA 777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAA 60

Qy        778 CGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATG 837
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATG 120

Qy        838 GCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGT 897
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGT 180

Qy        898 TGTAATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAA 957
                |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGTAATAAAAATATGATTAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAA 240

Qy        958 AATATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
                |||||||||||||||||| ||||||||||||||||||||||||
Db        241 AATATTACCGTGCTTAAGAATCTTACTAAACGGATAATTAAGTAG 285


RESULT 11
AY062288
LOCUS       AY062288               1832 bp    DNA     linear   BCT 20-MAR-2002
DEFINITION  Salmonella enterica subsp. enterica serovar Travis wbaV-wbaU
            intergenic region, partial sequence.
ACCESSION   AY062288
VERSION     AY062288.1  GI:19550992
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Travis
  ORGANISM  Salmonella enterica subsp. enterica serovar Travis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 1832)
  AUTHORS   Wang,L., Andrianopoulos,K., Liu,D., Popoff,M.Y. and Reeves,P.R.
  TITLE     Extensive variation in the O-antigen gene cluster within one
            Salmonella enterica serogroup reveals an unexpected complex history
  JOURNAL   J. Bacteriol. 184 (6), 1669-1677 (2002)
   PUBMED   11872718
REFERENCE   2  (bases 1 to 1832)
  AUTHORS   Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2001) Microbiology, The University of Sydney,
            Sydney, NSW 2006, Australia
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Salmonella enterica subsp. enterica serovar
                     Travis"
                     /mol_type="genomic DNA"
                     /serovar="Travis"
                     /sub_species="enterica"
                     /db_xref="taxon:181830"
                     /note="O-antigen gene cluster"
     misc_feature    <1. .>1832
                     /note="wbaV-wbaU intergenic region"
ORIGIN

  Query Match             27.7%;  Score 277.8;  DB 1;  Length 1832;
  Best Local Similarity   99.3%;  Pred. No. 7.3e-37;
  Matches  279;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        722 TTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAG 60

Qy        782 CATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCAT 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCAT 120

Qy        842 GTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        121 GTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGCA 180

Qy        902 ATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATA 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATA 240

Qy        962 TTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
              |||||||||||||||| ||||||||||||||||||||||||
Db        241 TTACCGTGCTTAAGAATCTTACTAAACGGATAATTAAGTAG 281


RESULT 12
AY062290
LOCUS       AY062290               1832 bp    DNA     linear   BCT 20-MAR-2002
DEFINITION  Salmonella enterica subsp. enterica serovar Tennyson wbaV-wbaU
            intergenic region, partial sequence.
ACCESSION   AY062290
VERSION     AY062290.1  GI:19550994
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Tennyson
  ORGANISM  Salmonella enterica subsp. enterica serovar Tennyson
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 1832)
  AUTHORS   Wang,L., Andrianopoulos,K., Liu,D., Popoff,M.Y. and Reeves,P.R.
  TITLE     Extensive variation in the O-antigen gene cluster within one
            Salmonella enterica serogroup reveals an unexpected complex history
  JOURNAL   J. Bacteriol. 184 (6), 1669-1677 (2002)
   PUBMED   11872718
REFERENCE   2  (bases 1 to 1832)
  AUTHORS   Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2001) Microbiology, The University of Sydney,
            Sydney, NSW 2006, Australia
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Salmonella enterica subsp. enterica serovar
                     Tennyson"
                     /mol_type="genomic DNA"
                     /serovar="Tennyson"
                     /sub_species="enterica"
                     /db_xref="taxon:181829"
                     /note="O-antigen gene cluster"
     misc_feature    <1. .>1832
                     /note="wbaV-wbaU intergenic region"
ORIGIN

  Query Match             27.7%;  Score 277.8;  DB 1;  Length 1832;
  Best Local Similarity   99.3%;  Pred. No. 7.3e-37;
  Matches  279;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        722 TTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAG 60

Qy        782 CATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCAT 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCAT 120

Qy        842 GTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        121 GTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGCA 180

Qy        902 ATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATA 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATA 240

Qy        962 TTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
              |||||||||||||||| ||||||||||||||||||||||||
Db        241 TTACCGTGCTTAAGAATCTTACTAAACGGATAATTAAGTAG 281


RESULT 13
AY062291
LOCUS       AY062291               1832 bp    DNA     linear   BCT 20-MAR-2002
DEFINITION  Salmonella enterica subsp. enterica serovar Texas wbaV-wbaU
            intergenic region, partial sequence.
ACCESSION   AY062291
VERSION     AY062291.1  GI:19550995
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Texas
  ORGANISM  Salmonella enterica subsp. enterica serovar Texas
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
```

```
REFERENCE   1  (bases 1 to 1832)
  AUTHORS   Wang,L., Andrianopoulos,K., Liu,D., Popoff,M.Y. and Reeves,P.R.
  TITLE     Extensive variation in the O-antigen gene cluster within one
            Salmonella enterica serogroup reveals an unexpected complex history
  JOURNAL   J. Bacteriol. 184 (6), 1669-1677 (2002)
   PUBMED   11872718
REFERENCE   2  (bases 1 to 1832)
  AUTHORS   Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2001) Microbiology, The University of Sydney,
            Sydney, NSW 2006, Australia
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Salmonella enterica subsp. enterica serovar
                     Texas"
                     /mol_type="genomic DNA"
                     /serovar="Texas"
                     /sub_species="enterica"
                     /db_xref="taxon:181737"
                     /note="O-antigen gene cluster"
     misc_feature    <1. .>1832
                     /note="wbaV-wbaU intergenic region"
ORIGIN

  Query Match             27.7%;  Score 277.8;  DB 1;  Length 1832;
  Best Local Similarity   99.3%;  Pred. No. 7.3e-37;
  Matches  279;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        722 TTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAG 60

Qy        782 CATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCAT 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCAT 120

Qy        842 GTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        121 GTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGCA 180

Qy        902 ATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATA 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATA 240

Qy        962 TTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
              |||||||||||||||| ||||||||||||||||||||||
Db        241 TTACCGTGCTTAAGAATCTTACTAAACGGATAATTAAGTAG 281


RESULT 14
AY062292
LOCUS       AY062292               1832 bp    DNA     linear   BCT 20-MAR-2002
DEFINITION  Salmonella enterica subsp. enterica serovar Azteca wbaV-wbaU
            intergenic region, partial sequence.
ACCESSION   AY062292
VERSION     AY062292.1  GI:19550996
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Azteca
  ORGANISM  Salmonella enterica subsp. enterica serovar Azteca
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 1832)
  AUTHORS   Wang,L., Andrianopoulos,K., Liu,D., Popoff,M.Y. and Reeves,P.R.
  TITLE     Extensive variation in the O-antigen gene cluster within one
            Salmonella enterica serogroup reveals an unexpected complex history
  JOURNAL   J. Bacteriol. 184 (6), 1669-1677 (2002)
   PUBMED   11872718
REFERENCE   2  (bases 1 to 1832)
  AUTHORS   Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2001) Microbiology, The University of Sydney,
            Sydney, NSW 2006, Australia
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Salmonella enterica subsp. enterica serovar
                     Azteca"
                     /mol_type="genomic DNA"
                     /serovar="Azteca"
                     /sub_species="enterica"
                     /db_xref="taxon:181735"
                     /note="O-antigen gene cluster"
     misc_feature    <1. .>1832
                     /note="wbaV-wbaU intergenic region"
ORIGIN

  Query Match             27.7%;  Score 277.8;  DB 1;  Length 1832;
  Best Local Similarity   99.3%;  Pred. No. 7.3e-37;
  Matches  279;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        722 TTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAG 60

Qy        782 CATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCAT 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCAT 120

Qy        842 GTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        121 GTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGCA 180

Qy        902 ATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATA 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATA 240

Qy        962 TTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
              |||||||||||||||| ||||||||||||||||||||||
Db        241 TTACCGTGCTTAAGAATCTTACTAAACGGATAATTAAGTAG 281


RESULT 15
AY062289
LOCUS       AY062289               1834 bp    DNA     linear   BCT 20-MAR-2002
DEFINITION  Salmonella enterica subsp. enterica serovar Tejas wbaV-wbaU
            intergenic region, partial sequence.
ACCESSION   AY062289
VERSION     AY062289.1  GI:19550993
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Tejas
  ORGANISM  Salmonella enterica subsp. enterica serovar Tejas
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 1834)
  AUTHORS   Wang,L., Andrianopoulos,K., Liu,D., Popoff,M.Y. and Reeves,P.R.
  TITLE     Extensive variation in the O-antigen gene cluster within one
            Salmonella enterica serogroup reveals an unexpected complex history
  JOURNAL   J. Bacteriol. 184 (6), 1669-1677 (2002)
   PUBMED   11872718
REFERENCE   2  (bases 1 to 1834)
  AUTHORS   Wang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2001) Microbiology, The University of Sydney,
            Sydney, NSW 2006, Australia
FEATURES             Location/Qualifiers
     source          1. .1834
                     /organism="Salmonella enterica subsp. enterica serovar
                     Tejas"
                     /mol_type="genomic DNA"
                     /serovar="Tejas"
                     /sub_species="enterica"
                     /db_xref="taxon:181828"
                     /note="O-antigen gene cluster"
     misc_feature    <1. .>1834
                     /note="wbaV-wbaU intergenic region"
ORIGIN
```

```
  Query Match             27.7%;  Score 277.8;  DB 1;  Length 1834;
  Best Local Similarity   99.3%;  Pred. No. 7.3e-37;
  Matches  279;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        722 TTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAG 60

Qy        782 CATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCAT 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCAT 120

Qy        842 GTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        121 GTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGCA 180

Qy        902 ATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATA 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATA 240

Qy        962 TTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
              |||||||||||||||| ||||||||||||||||||||||
Db        241 TTACCGTGCTTAAGAATCTTACTAAACGGATAATTAAGTAG 281


Search completed: May 11, 2006, 19:15:07
Job time : 3610.66 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 07:09:59 ; Search time 456.564 Seconds
(without alignments)
14626.705 Million cell updates/sec

Title: US-09-423-093B-4_COPY_14059_15060
Perfect score: 1002
Sequence: 1 atgcttatatcattttgtat.....ctaaacggataattaagtag 1002

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1002	100.0	22080	2	AAX06751	Aax06751 Salmonell
	2	50.2	5.0	73334	6	ABL34124	Abl34124 Human imm
	3	50.2	5.0	73334	6	ABL92318	Abl92318 Chemicall
c	4	50.2	5.0	99764	13	ABD32992	Abd32992 Human can
	5	49.6	5.0	894	8	ACA39370	Aca39370 Prokaryot
	6	49.6	5.0	2601	10	AAD63519	Aad63519 Mycoplasm
	7	49.6	5.0	2601	10	ACC69149	Acc69149 M. genita
	8	49.6	5.0	2601	12	ADN48954	Adn48954 Mycoplasm
	9	49.6	5.0	3480	2	AAQ55143	Aaq55143 Enterococ
	10	49.6	5.0	3480	8	ABZ77359	Abz77359 Nucleotid
	11	49.6	5.0	3480	10	AAL51847	Aal51847 Enterococ
	12	49.6	5.0	110000	2	AAT58840_0	Aat58840 Mycoplasm
c	13	49.2	4.9	2012	2	AAX13380	Aax13380 Enterococ
c	14	49.2	4.9	2012	6	ABS99175	Abs99175 Enterococ
	15	48	4.8	477	10	ADH82927	Adh82927 Enterococ
c	16	47.8	4.8	2000	11	ACL37108	Acl37108 Rice stre
	17	47.4	4.7	110000	6	ABA92787_3	Continuation (4 of
	18	47	4.7	10048	6	ABQ67016	Abq67016 Human ang
c	19	46.8	4.7	3001	3	AAH51792	Aah51792 Chromosom
c	20	46.6	4.7	2000	11	ACL35887	Acl35887 Rice stre
	21	46.6	4.7	2431	6	ABL34268	Abl34268 Human imm
	22	46.6	4.7	18997	6	ABL32571	Abl32571 Human imm
	23	46.6	4.7	18997	6	ABK33949	Abk33949 Human DNA
	24	46.6	4.7	18997	8	ADA20353	Ada20353 Prostate
	25	46.6	4.7	18997	8	ADA84160	Ada84160 Human ren
	26	46.6	4.7	19787	6	ABL33451	Abl33451 Human imm
c	27	46.6	4.7	110000	6	ABA92787_2	Continuation (3 of
c	28	46.6	4.7	110000	13	ABD32968_6	Continuation (7 of
	29	46.4	4.6	9997	6	ABL32369	Abl32369 Human imm
	30	46.4	4.6	9997	6	ABL34485	Abl34485 Human met
	31	46.4	4.6	9997	7	ADS99746	Ads99746 Complemen
	32	46.4	4.6	12441	2	AAX06750	Aax06750 Salmonell
	33	46.2	4.6	110000	13	ABD32629_1	Continuation (2 of
	34	46	4.6	543	13	ACN57427	Acn57427 Cotton gy
c	35	46	4.6	9300	5	ABA15594	Aba15594 Human ner
c	36	46	4.6	65359	6	AAD46665	Aad46665 Human tra
	37	46	4.6	337344	13	ABD32715	Abd32715 Human can
c	38	45.6	4.6	8056	8	ABZ10246	Abz10246 Haematopo
	39	45.4	4.5	2000	11	ACL35887	Acl35887 Rice stre
	40	45.4	4.5	34688	6	ABQ67060	Abq67060 Human ang
	41	45.4	4.5	73334	6	ABL34125	Abl34125 Human imm
	42	45.4	4.5	73334	6	ABL92319	Abl92319 Chemicall
	43	45	4.5	1152	6	ABZ31993	Abz31993 Candida a
c	44	45	4.5	2000	11	ACL35363	Acl35363 Rice stre
	45	45	4.5	5768	6	ABL34163	Abl34163 Human imm

ALIGNMENTS

RESULT 1
AAX06751
ID AAX06751 standard; DNA; 22080 BP.
XX
AC AAX06751;
XX
DT 26-APR-1999 (first entry)
XX
DE Salmonella enterica serovar B O antigen gene cluster.
XX
KW O antigen; diagnosis; wzx gene; wbaV gene; transferase; polymerase;
KW flippase; diarrhoea; haemorrhagic colitis; ds.
XX
OS Salmonella enterica.
XX

FH	Key	Location/Qualifiers
FT	CDS	4099. .5184
FT		/*tag= a
FT		/product= "TDP-rhamnose_pathway_protein"
FT		/note= "rmlB gene; encodes AAW88335"
FT	CDS	5184. .6083
FT		/*tag= b
FT		/product= "TDP-rhamnose_pathway_protein"
FT		/note= "rmlD gene; encodes AAW88336"
FT	CDS	6131. .7009
FT		/*tag= c
FT		/product= "TDP-rhamnose_pathway_protein"
FT		/note= "rmlA gene; encodes AAW88337"
FT	CDS	7010. .7561
FT		/*tag= d
FT		/product= "TDP-rhamnose_pathway_protein"
FT		/note= "rmlC gene; encodes AAW88338"
FT	CDS	7567. .8559
FT		/*tag= e
FT		/product= "CDP-adequose_pathway_protein"
FT		/note= "ddhD gene; encodes AAW88339"
FT	CDS	8556. .9329
FT		/*tag= f
FT		/product= "CDP-adequose_pathway_protein"
FT		/note= "ddhA gene; encodes AAW88340"
FT	CDS	9334. .10413
FT		/*tag= g

```
FT                   /product= "CDP-adequose_pathway_protein"
FT                   /note= "ddhB gene; encodes AAW88341"
FT   CDS             10440. .11753
FT                   /*tag= h
FT                   /product= "CDP-adequose_pathway_protein"
FT                   /note= "ddhC gene; encodes AAW88342"
FT   CDS             11781. .12680
FT                   /*tag= i
FT                   /product= "CDP-adequose_pathway_protein"
FT                   /note= "abe gene; encodes AAW88343"
FT   CDS             12762. .14054
FT                   /*tag= j
FT                   /product= "flippase"
FT                   /note= "wzx gene, this region is specifically claimed in
FT                   Claim 9; encodes AAW88344"
FT   CDS             14059. .15060
FT                   /*tag= k
FT                   /product= "abequosyl_transferase"
FT                   /note= "wbaV gene, this region is specifically claimed in
FT                   Claim 9; ; encodes AAW88345"
FT   CDS             15379. .16440
FT                   /*tag= l
FT                   /product= "mannosyl_transferase"
FT                   /note= "wbaU gene; encodes AAW88346"
FT   CDS             16441. .17385
FT                   /*tag= m
FT                   /product= "rhamnosyl_transferase"
FT                   /note= "wbaN gene; encodes AAW88347"
FT   CDS             17386. .18825
FT                   /*tag= n
FT                   /product= "GDP_mannose_pathway_protein"
FT                   /note= "manC gene; encodes AAW88348"
FT   CDS             18812. .20245
FT                   /*tag= o
FT                   /product= "GDP_mannose_pathway_protein"
FT                   /note= "mabB gene; encodes AAW88349"
FT   CDS             20317. .21747
FT                   /*tag= h
FT                   /product= "galactosyl_transferase"
FT                   /note= "wbaP gene; encodes AAW88350"
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU000315.
XX
PR   01-MAY-1997;   97AU-00006545.
PR   22-JUL-1997;   97AU-00008162.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   P-PSDB; AAW88335, AAW88336, AAW88337, AAW88338, AAW88339, AAW88340,
DR   AAW88341, AAW88342, AAW88343, AAW88344, AAW88345, AAW88346, AAW88347,
DR   AAW88348, AAW88349, AAW88350.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide antigens -
PT   useful for detecting specific strains in, e.g. food, faeces or patient
PT   samples.
XX
PS   Claim 9; Fig 10; 165pp; English.
XX
CC   This is the nucleotide sequence of a gene cluster involved in the
CC   biosynthesis of the Salmonella enterica serotype B (strain LT2) O
CC   antigen. 16 Open reading frames within the gene cluster were identified
CC   on the basis of homology to known sequences. The genes encode proteins
CC   (see AAW88335-50) involved the synthesis of sugars present in the
CC   polysaccharide antigen, and in the transport or processing of
CC   polysaccharide or oligosaccharide units. The use of nucleic acid
CC   molecules derived from assembly and transport genes, particularly wbd
CC   (transferase), wzx (flippase) and wzy (polymerase) genes, within O
CC   antigen gene clusters improves the specificity of methods for the
CC   detection and identification of O antigens, e.g. in testing food- or
CC   faecal-derived samples, or samples from patients. The O antigen is a
CC   major virulence factor of S. enterica
XX
SQ   Sequence 22080 BP; 6284 A; 4028 C; 5152 G; 6616 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 1002;  DB 2;  Length 22080;
  Best Local Similarity  100.0%;  Pred. No. 5.2e-205;
  Matches 1002;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14059 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 14118

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14119 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 14178

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14179 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 14238

Qy        181 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14239 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 14298

Qy        241 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14299 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 14358

Qy        301 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14359 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 14418

Qy        361 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14419 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 14478

Qy        421 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14479 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 14538

Qy        481 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14539 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 14598

Qy        541 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14599 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 14658

Qy        601 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14659 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 14718

Qy        661 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14719 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 14778

Qy        721 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14779 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 14838

Qy        781 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14839 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 14898

Qy        841 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14899 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 14958

Qy        901 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 960
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14959 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 15018

Qy        961 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
              |||||||||||||||||||||||||||||||||||||||||
Db      15019 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 15060


RESULT 2
ABL34124
ID   ABL34124 standard; DNA; 73334 BP.
XX
AC   ABL34124;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2097.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 2097; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 73334 BP; 18968 A; 858 C; 15329 G; 38179 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 50.2;  DB 6;  Length 73334;
  Best Local Similarity   45.2%;  Pred. No. 0.76;
  Matches  184;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy        474 TAGATGCTTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGA 533
              |||    |  | ||| |  ||||| ||||| ||  | | |   | |  | ||||  |
Db      36186 TAGTGTTTGTTGTATAGTAGGTGTTTTTAGGTAAATGATTGAATGATAAATAAATTATAT 36245

Qy        534 ACGATGGGATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATT 593
              | |  |||   ||| |     |       | | ||     ||  |        |||
Db      36246 CGGTTATGATAGTATTTAAAAGAAAATAGTTTTTTTTTTGTTTTTTTTGTAAGGAGTAAG 36305

Qy        594 TATCATGATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACT 653
              |  |  |   | | ||  ||    ||     ||  | |||   |   | |   |
Db      36306 GAAATTTTTTTTGGAAGTTTTTATATTAGATATTTTTACGTATTGGTTTAGGATATAGTT 36365

Qy        654 CGTAATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTT 713
              | ||| |   | |  ||  |  |||||       |||||  |      |   |||| |
Db      36366 ATTGATAAGGGAATGGAATTTTTTTAGTTGGTTTGGAAAAGGTAGTTTGTGTTTAATTAT 36425

Qy        714 AATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTT 773
              ||    | || |       | ||||| ||    ||||||  ||| || |||  ||||  ||
Db      36426 AAAGATTATTGTATTGGTTATAAAAATAATATATGATTATTATAGGAAATTTATAAATTA 36485

Qy        774 AAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGC 833
              |  |  || ||  |||||  |||| ||  | | | |  |   |  | | |  |  || |
Db      36486 ATGAAAAGTATAATAAAATAATTTGGTATTATATAAAGAAAATTGTTGTTAATATTGTGG 36545

Qy        834 TATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAAT 880
              | |   | ||  |   | ||  |  |   | |  ||| | || ||
Db      36546 TGTTTGTTTTTTAGTTTTTGGTTGTATATGTATATGTTTGTGTGTAT 36592


RESULT 3
ABL92318
ID   ABL92318 standard; DNA; 73334 BP.
XX
AC   ABL92318;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Chemically treated DNA repair gene fragment#64.
XX
KW   DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L2; PMS2L3; PMS2;
KW   L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4; DDIT1L; FANCB;
KW   XRCC8; ataxia telangiectasia; aging; Bloom's syndrome; Cockayne syndrome;
KW   Nijmegen breakage syndrome; Werner syndrome; immunodeficiency;
KW   trichthiodystrophy; Fanconi's anaemia; solid tumour; cancer; ds.
XX
OS   Unidentified.
XX
PN   WO200181622-A2.
XX
PD   01-NOV-2001.
XX
PF   06-APR-2001; 2001WO-EP003972.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-034446/04.
XX
PT   New nucleic acid derived from genes associated with DNA repair, useful
PT   for diagnosis, e.g. of ataxia telangiectasia, by determination of
PT   cytosine methylation.
XX
PS   Claim 1; SEQ ID NO 127; 25pp + Sequence Listing; English.
XX
CC   The invention relates to nucleic acids containing a sequence of at least
CC   18 nucleotides of chemically treated DNA of genes associated with DNA
CC   repair, and their complements. The invention also relates to nucleic
CC   acids comprising at least 18 base pairs of the chemically pretreated DNA
CC   of genes associated with DNA repair selected from PMS2L1, PMS2L12,
CC   PMS2L2, PMS2L3, PMS2, L4, PMS2L5, PMS2L6, MGMT, MSH2, NUDT1, TDG, INPPL1,
CC   RFC4, DDIT1L, FANCB, or XRCC8. Nucleic acids of the invention and related
CC   oligomers, are useful for diagnosis of diseases associated with gene
CC   repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
CC   Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,
CC   immunodeficiency, trichthiodystrophy, Fanconi's anaemia, solid tumours
```

CC and cancer, particularly by determining status of cytosine methylation
CC and/or by detecting single-nucleotide polymorphisms. Determination of
CC individual methylation patterns may allow development of individualised
CC therapies. The sequences given in records ABL92192-ABL92335 represent
CC chemically pre-treated DNA fragments from genes associated with DNA
CC repair, and their complements. Note: The sequence data for this patent is
CC not represented in the specification, but is based on sequence
CC information supplied by the European Patent Office
XX
SQ Sequence 73334 BP; 18968 A; 858 C; 15329 G; 38179 T; 0 U; 0 Other;

Query Match 5.0%; Score 50.2; DB 6; Length 73334;
Best Local Similarity 45.2%; Pred. No. 0.76;
Matches 184; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

Qy 474 TAGATGCTTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGA 533
Db 36186 TAGTGTTTGTTGTATAGTAGGTGTTTTTAGGTAAATGATTGAATGATAAATAAATTATAT 36245

Qy 534 ACGATGGGATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATT 593
Db 36246 CGGTTATGATAGTATTTAAAAGAAAATAGTTTTTTTTTTGTTTTTTTTGTAAGGAGTAAG 36305

Qy 594 TATCATGATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACT 653
Db 36306 GAAATTTTTTTTGGAAGTTTTATATTAGATATTTTTACGTATTGGTTTAGGATATAGTT 36365

Qy 654 CGTAATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTT 713
Db 36366 ATTGATAAGGGAATGGAATTTTTTTAGTTGGTTTGGAAAAGGTAGTTTGTGTTTAATTAT 36425

Qy 714 AATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTT 773
Db 36426 AAAGATTATTGTATTGGTTATAAAAATAATATATGATTATTATAGGAAATTTATAAATTA 36485

Qy 774 AAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGC 833
Db 36486 ATGAAAAGTATAATAAAATAATTTGGTATTATATAAAGAAAATTGTTGTTAATATTGTGG 36545

Qy 834 TATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAAT 880
Db 36546 TGTTTGTTTTTTAGTTTTTGGTTGTATATGTATATGTTTGTGTGTAT 36592

RESULT 4
ABD32992/c
ID ABD32992 standard; DNA; 99764 BP.
XX
AC ABD32992;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated genomic DNA HD21-017.
XX
KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW leukaemia; lymphoma; CAP.
XX
OS Homo sapiens.
XX
PN WO2004074320-A2.
XX
PD 02-SEP-2004.
XX
PF 17-FEB-2004; 2004WO-US004730.
XX
PR 14-FEB-2003; 2003US-00367094.
PR 14-MAR-2003; 2003US-00388838.
PR 15-APR-2003; 2003US-00417375.
PR 13-JUN-2003; 2003US-00461862.
PR 15-SEP-2003; 2003US-00663431.
PR 15-DEC-2003; 2003US-00737318.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Morris DW, Malandro MS;
XX
DR WPI; 2004-652914/63.
XX
PT New isolated cancer-associated polynucleotides and polypeptides useful
PT for diagnosing, preventing or treating cancers, especially lymphoma and
PT leukemia, or in screening for agents that modulate cancer.
XX
PS claim 16; seqid 754; 310pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising at least 10
CC contiguous nucleotides of any of the 233 polynucleotide sequences given
CC in the specification, or its complement. The nucleic acids encode cancer-
CC associated proteins. Also included are an expression vector comprising
CC the isolated nucleic acid cited above, a host cell comprising the above
CC recombinant nucleic acid or expression vector, a microarray for detecting
CC a cancer-associated (CA) nucleic acid comprising at least one probe
CC comprising at least 10 contiguous nucleotides of any of the above-
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC an open reading frame of a CA sequence selected from any of the 95
CC polynucleotide sequences as mentioned in the specification, or its
CC complement), an isolated antibody, (or its antigen binding fragment) that
CC binds to the above polypeptide, a hybridoma that produces the above
CC monoclonal antibody, a pharmaceutical composition comprising the above
CC antibody and a pharmaceutical excipient, a kit for detecting cancer
CC cells(comprising the antibody cited above, methods for diagnosing cancer
CC or for detecting the presence or absence of cancer cells in an
CC individual, a method for inhibiting growth of cancer cells in an
CC individual, a method for delivering a therapeutic agent to cancer cells
CC in an individual, an electronic library comprising the above
CC polynucleotide or polypeptide (or their fragments), methods of screening
CC for anticancer activity or for a bioactive agent capable of modulating
CC the activity of a CA protein (CAP), methods for detecting cancer
CC associated with expression of a polypeptide in a test cell sample, a
CC method for treating cancers and a method for inhibiting the expression of
CC CA gene in a cell. The composition and methods are useful for detecting,
CC diagnosing, preventing and treating cancers, especially lymphoma and
CC leukaemia. These may also be used in screening for agents that modulate
CC cancer. The present sequence is a human CAP genomic sequence. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 99764 BP; 29560 A; 17120 C; 18480 G; 34604 T; 0 U; 0 Other;

Query Match 5.0%; Score 50.2; DB 13; Length 99764;
Best Local Similarity 44.0%; Pred. No. 0.79;
Matches 259; Conservative 0; Mismatches 328; Indels 2; Gaps 1;

Qy 372 GTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACAGATGATGAACT 431
Db 41512 GTTATATTATATTTTGTATATAATATATAATTGTTATATTTTGTATATAATATATAATTA 41453

Qy 432 TTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATCTATTGG 491
Db 41452 CATATTATATATTAATAATTATTATATATTAATATATAATATATAATAATTATTAATATA 41393

Qy 492 TGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGGGATGCCATTGA 551
Db 41392 TAATATATTAATATATAATAAATATATATTAATAATTAAAATATAATATAATTCTATATT 41333

Qy 552 TTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGCGTATT 611
Db 41332 ATATTTTGTATAATATAAAATCATATATTATATATAATTCTATATTATGTTATATTTTGT 41273

Qy 612 TAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATATGCCGAGGAGA 671
Db 41272 ATATAATATATAATCATATATAATATATAATTCTATATTATATTTTGTATAT--AATATA 41215

Qy 672 TAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATA 731
Db 41214 TAATTATATATTATATATAATTCTATATACTTTGTATATAATATATAATTCTGTTATATT 41155

```
Qy        732 TTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAA 791
              ||  | || |  ||    | | |    | | | | | | | ||| |       | ||  |
Db      41154 TTGTATATAATATACAATTCTGTTATATTACATTTTGTATATAATATACAATTCTGTTAT 41095

Qy        792 TGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCATGTTATGGCAA 851
                |||   || || | | |       |  ||||      |              |  |||
Db      41094 ATTTTGTATATAATATACAATTCAATGTTATATTTTGTATATACCATTGTTATATTTTGT 41035

Qy        852 TAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATAT 911
                | ||  | ||      |  | || || | ||  | |  |  ||| || || |  |
Db      41034 ATATAATATATAATTGTTATATTTTGTATATAATATATAATTCTGTTATATTATATTTTG 40975

Qy        912 GATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 960
               ||  |   | || |||  ||     |  |    |||  | |  | |||
Db      40974 TATATAATATATAATTCTGTTATATTATATTTTGTATATAATATATAAT 40926
```

RESULT 5
ACA39370
ID ACA39370 standard; DNA; 894 BP.
XX
AC ACA39370;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #21027.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Mycoplasma genitalium.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU35500.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 27240; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 894 BP; 333 A; 101 C; 143 G; 317 T; 0 U; 0 Other;

  Query Match   5.0%; Score 49.6; DB 8; Length 894;
  Best Local Similarity   46.0%; Pred. No. 0.51;
  Matches 209; Conservative   0; Mismatches 239; Indels   6; Gaps   1;

```
Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||    ||||  |   ||| || ||||| ||| |   |  || | |||||  |     |
Db          1 ATGAAATTATCTGTAATTATACCTACTTACAATTGTGCATCATTTATTGAAAAAGCAATT 60

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              |||   ||   |||  |   |     | |||| || || ||  | |  ||    ||| ||
Db         61 AATTCAATTGTTAAAAATAGACCTAATGATTTGGAAATAGAAGTTTTAATTATTGATGAT 120

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              |  || | |||   |||  |  || | ||| |            || |   ||||      |
Db        121 GGATCAATTGACAATACTAACAAAGTTATTAAGAAAATTCAAGACCAAATTAATAATTTA 180

Qy        181 ATAATATATCGGCGT------AATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCT 234
              |   |  |       |      | ||  | ||||   ||    | ||  |||| | ||
Db        181 ACTTTGCAGTATTTTTACAAAAGTAATGGTAACTGGGGTAGTGTTATTAATTATGTTAGA 240

Qy        235 TCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCG 294
                  || || || ||||| |          |||| |  |||||||    | || |
Db        241 AACAATAAACTAGCAAAAGGGGAATGAGTAACAGTATTGGATAGTGATGACATTTTTTCA 300

Qy        295 AAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGT 354
              ||| |  |  | | || || |||| |||  | |    |||  | ||       |  |
Db        301 AAAAAAACAATTTCTATTTTTCAAAAATATGCCCAAAAACAAAGATATGATGCGATTATT 360

Qy        355 GACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTC 414
                  | | |   | || |     || |   | ||  |||     | |  || | |
Db        361 TTTGACTACTATAAATGCTGAAAAAAGTTTTTGTGAAAAATTCCTACCTATGCAAGGTTT 420

Qy        415 AGAACAGATGATGAACTTTATGTGTTTAATAATA 448
              |||| |||   | ||  | |  ||   ||  | |
Db        421 AGAAAAGAAATTAAAGGTGAATTGAAAAAACAAA 454
```

RESULT 6
AAD63519
ID AAD63519 standard; DNA; 2601 BP.
XX
AC AAD63519;
XX
DT 12-FEB-2004 (first entry)
XX
DE Mycoplasma genitalium gene involved in particle envelope.
XX
KW Genetic operating system; nanomachine genome; bioreactor; bioremediation;
KW therapeutic biomolecule; energy conversion system; processing system;
KW anabolic; catabolic system; biological film; cosmetic application;
KW coating; ds.
XX
OS Mycoplasma genitalium.

XX
PN US2003138777-A1.
XX
PD 24-JUL-2003.
XX
PF 20-SEP-2001; 2001US-00960858.
XX
PR 20-SEP-2001; 2001US-00960858.
XX
PA (EVAN/) EVANS G A.
XX
PI Evans GA;
XX
DR WPI; 2003-851721/79.
XX
PT Basic genetic operating system for an autonomous prototrophic
PT nanomachine, comprises a nanomachine genome encoding a minimal gene set.
XX
PS Example 1; Page 151-152; 170pp; English.
XX
CC The invention relates to a basic genetic operating system which comprises
CC a nanomachine genome encoding a minimal gene set for viability. The basic
CC genetic operating system is used for an autonomous prototrophic
CC nanomachine or autonomous auxotrophic nanomachine. The nanomachine is
CC used as a bioreactor, for bioremediation, for production of a therapeutic
CC biomolecule or as therapeutic agent, for production of a diagnostic
CC indicator or as a diagnostic reagent, as a delivery system, as an
CC artificial tissue or organ system, an energy conversion system, as a
CC processing system, as an anabolic or catabolic system, for production of
CC biological films or coatings that may respond to the environment and for
CC cosmetic applications including pharmaceuticals. The present sequence is
CC Mycoplasma genitalium nanomachine gene
XX
SQ Sequence 2601 BP; 888 A; 333 C; 434 G; 946 T; 0 U; 0 Other;

Query Match 5.0%; Score 49.6; DB 10; Length 2601;
Best Local Similarity 46.0%; Pred. No. 0.6;
Matches 209; Conservative 0; Mismatches 239; Indels 6; Gaps 1;

```
Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||    ||||  |   ||| || ||||| ||| |   |  || | |||||  |     |
Db        519 ATGAAATTATCTGTAATTATACCTACTTACAATTGTGCATCATTTATTGAAAAAGCAATT 578

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              |||   ||   |||  |   |     | |||| || || ||  | |  ||    ||| ||
Db        579 AATTCAATTGTTAAAAATAGACCTAATGATTTGGAAATAGAAGTTTTAATTATTGATGAT 638

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              |  || | |||  |||  |  || | ||| |           || |   ||||     |
Db        639 GGATCAATTGACAATACTAACAAAGTTATTAAGAAAATTCAAGACCAAATTAATAATTTA 698

Qy        181 ATAATATATCGGCGT------AATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCT 234
              |   |  |       |      | ||  | ||||   ||    | ||  |||| | ||
Db        699 ACTTTGCAGTATTTTTACAAAAGTAATGGTAACTGGGGTAGTGTTATTAATTATGTTAGA 758

Qy        235 TCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCG 294
                 || || || ||||| |          ||||  |  ||||||||   | || |
Db        759 AACAATAAACTAGCAAAAGGGGAATGAGTAACAGTATTGGATAGTGATGACATTTTTTCA 818

Qy        295 AAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGT 354
              ||| |  |  |  | || || ||||  |||  | |    |||  | ||        |   |
Db        819 AAAAAAACAATTTCTATTTTTCAAAAATATGCCCAAAAACAAAGATATGATGCGATTATT 878

Qy        355 GACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTC 414
                | | |   | ||  |    ||  |  | || |||    | |  || |
Db        879 TTTGACTACTATAAATGCTGAAAAAAGTTTTTGTGAAAAATTCCTACCTATGCAAGGTTT 938

Qy        415 AGAACAGATGATGAACTTTATGTGTTTAATAATA 448
              |||| |||   | ||  | |  ||   ||  | |
Db        939 AGAAAAGAAATTAAAGGTGAATTGAAAAAACAAA 972
```

RESULT 7

ACC69149
ID ACC69149 standard; DNA; 2601 BP.
XX
AC ACC69149;
XX
DT 10-JUL-2003 (first entry)
XX
DE M. genitalium particle envelope gene cassette DNA SEQ ID NO:17.
XX
KW Mycoplasma genitalium; gene cassette; replication; transcription;
KW translation; metabolism; basic genetic operating system; gene therapy;
KW autonomous prototrophic nanomachine; auxotrophic nanomachine;
KW nanomachine; bioreactor; bioremediation; therapeutic; delivery system;
KW artificial tissue; artificial organ system; energy conversion system;
KW processing system; anabolic system; catabolic system; biological film;
KW biological coating; cosmetic; gene; ds.
XX
OS Mycoplasma genitalium.
XX
PN WO2003025145-A2.
XX
PD 27-MAR-2003.
XX
PF 18-SEP-2002; 2002WO-US029811.
XX
PR 20-SEP-2001; 2001US-00960870.
XX
PA (EGEA-) EGEA BIOSCIENCES INC.
XX
PI Evans GA;
XX
DR WPI; 2003-354602/33.
XX
PT New basic genetic operating system for autonomous prototrophic or
PT auxotrophic nanomachine, useful for therapeutic, diagnostic or industrial
PT purposes, comprises a nanomachine genome encoding a gene set for
PT viability or replication.
XX
PS Example 1; Page 247-248; 250pp; English.
XX
CC The present invention describes a basic genetic operating system for an
CC autonomous prototrophic or auxotrophic nanomachine comprising a
CC nanomachine genome encoding a minimal gene set sufficient for viability
CC or replication, optionally in the presence of an auxotrophic molecule.
CC Also described is an autonomous prototrophic or auxotrophic nanomachine
CC comprising a basic genetic operating system for autonomous prototrophic
CC or auxotrophic viability or replication, optionally in the presence of an
CC auxotrophic molecule, and a particle envelope. The nanomachines can be
CC used in gene therapy. The basic genetic operating system or nanomachine
CC is useful in therapeutic, diagnostic and industrial applications, e.g. as
CC a bioreactor, for bioremediation, for the production of a therapeutic
CC biomolecule or as a therapeutic reagent, for the production of a
CC diagnostic indicator or reagent, as a delivery system, as an artificial
CC tissue or organ system, as an energy conversion system, as a processing
CC system, as an anabolic or catabolic system, for the production of
CC biological films or coatings, and for cosmetic applications. The present
CC sequence represents a Mycoplasma genitalium gene cassette nucleotide
CC sequence, which is used in an example from the present invention for the
CC design and synthesis of a basic genetic operation system for a
CC replication competent nanomachine
XX
SQ Sequence 2601 BP; 888 A; 333 C; 434 G; 946 T; 0 U; 0 Other;

Query Match 5.0%; Score 49.6; DB 10; Length 2601;
Best Local Similarity 46.0%; Pred. No. 0.6;
Matches 209; Conservative 0; Mismatches 239; Indels 6; Gaps 1;

```
Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||    ||||  |   ||| || ||||| ||| |   |  || | |||||  |     |
Db        519 ATGAAATTATCTGTAATTATACCTACTTACAATTGTGCATCATTTATTGAAAAAGCAATT 578

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              |||   ||   |||  |   |     | |||| || || ||  | |  ||    ||| ||
```

```
Db        579 AATTCAATTGTTAAAAATAGACCTAATGATTTGGAAATAGAAGTTTTAATTATTGATGAT 638

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180

Db        639 GGATCAATTGACAATACTAACAAAGTTATTAAGAAAATTCAAGACCAAATTAATAATTTA 698

Qy        181 ATAATATATCGGCGT------AATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCT 234

Db        699 ACTTTGCAGTATTTTTACAAAAGTAATGGTAACTGGGGTAGTGTTATTAATTATGTTAGA 758

Qy        235 TCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCG 294

Db        759 AACAATAAACTAGCAAAAGGGGAATGAGTAACAGTATTGGATAGTGATGACATTTTTTCA 818

Qy        295 AAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGT 354

Db        819 AAAAAAACAATTTCTATTTTTCAAAAATATGCCCAAAAACAAAGATATGATGCGATTATT 878

Qy        355 GACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTC 414

Db        879 TTTGACTACTATAAATGCTGAAAAAAGTTTTTGTGAAAAATTCCTACCTATGCAAGGTTT 938

Qy        415 AGAACAGATGATGAACTTTATGTGTTTAATAATA 448

Db        939 AGAAAAGAAATTAAAGGTGAATTGAAAAAACAAA 972
```

RESULT 8
ADN48954
ID ADN48954 standard; DNA; 2601 BP.
XX
AC ADN48954;
XX
DT 01-JUL-2004 (first entry)
XX
DE Mycoplasma genitalium functional gene fragment #17.
XX
KW Prototrophic nanomachine; functional gene; ds.
XX
OS Mycoplasma genitalium.
XX
PN US2004063097-A1.
XX
PD 01-APR-2004.
XX
PF 20-SEP-2002; 2002US-00251668.
XX
PR 20-SEP-2002; 2002US-00251668.
XX
PA (EVAN/) EVANS G A.
XX
PI Evans GA;
XX
DR WPI; 2004-346224/32.
XX
PT Basic genetic operating system for autonomous prototrophic nanomachine,
PT comprises nanomachine genome encoding minimal gene set sufficient for
PT viability.
XX
PS Example 1; SEQ ID NO 17; 170pp; English.
XX
CC The present invention provides a basic genetic operating system for an
CC autonomous prototrophic nanomachine comprises a nanomachine genome
CC encoding a minimal gene set sufficient for viability. The present
CC sequence is Mycoplasma genitalium functional gene fragment. This sequence
CC is used in the exemplification of the invention.
XX
SQ Sequence 2601 BP; 888 A; 333 C; 434 G; 946 T; 0 U; 0 Other;

Query Match 5.0%; Score 49.6; DB 12; Length 2601;
Best Local Similarity 46.0%; Pred. No. 0.6;
Matches 209; Conservative 0; Mismatches 239; Indels 6; Gaps 1;

```
Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60

Db        519 ATGAAATTATCTGTAATTATACCTACTTACAATTGTGCATCATTTATTGAAAAAGCAATT 578

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120

Db        579 AATTCAATTGTTAAAAATAGACCTAATGATTTGGAAATAGAAGTTTTAATTATTGATGAT 638

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180

Db        639 GGATCAATTGACAATACTAACAAAGTTATTAAGAAAATTCAAGACCAAATTAATAATTTA 698

Qy        181 ATAATATATCGGCGT------AATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCT 234

Db        699 ACTTTGCAGTATTTTTACAAAAGTAATGGTAACTGGGGTAGTGTTATTAATTATGTTAGA 758

Qy        235 TCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCG 294

Db        759 AACAATAAACTAGCAAAAGGGGAATGAGTAACAGTATTGGATAGTGATGACATTTTTTCA 818

Qy        295 AAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGT 354

Db        819 AAAAAAACAATTTCTATTTTTCAAAAATATGCCCAAAAACAAAGATATGATGCGATTATT 878

Qy        355 GACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTC 414

Db        879 TTTGACTACTATAAATGCTGAAAAAAGTTTTTGTGAAAAATTCCTACCTATGCAAGGTTT 938

Qy        415 AGAACAGATGATGAACTTTATGTGTTTAATAATA 448

Db        939 AGAAAAGAAATTAAAGGTGAATTGAAAAAACAAA 972
```

RESULT 9
AAQ55143
ID AAQ55143 standard; DNA; 3480 BP.
XX
AC AAQ55143;
XX
DT 25-MAR-2003 (revised)
DT 22-SEP-1995 (first entry)
XX
DE Enterococcus faecalis probe S2-7.
XX
KW Probe; S.aureus; S.epidermis; E.faecalis; P.aeruginosa; E.coli;
KW K.pneumoniae; E.cloacae; clinical sample; ds.
XX
OS Enterococcus faecalis.
XX
PN WO9401583-A1.
XX
PD 20-JAN-1994.
XX
PF 07-JUL-1993; 93WO-JP000936.
XX
PR 07-JUL-1992; 92JP-00179719.
XX
PA (FUSO) FUSO PHARM IND LTD.
XX
PI Ohno T, Matsuhisa A, Uehara H, Eda S;
XX
DR WPI; 1994-035086/04.
XX
PT Probe for identifying bacteria causing infectious disease - consists of a
PT DNA fragment obtained by HindIII cleavage of the pathogenic bacterial
PT genomic DNA.
XX
PS Claim 5; Page 45-47; 100pp; Japanese.
XX
CC The nucleotide sequence of a 3480 bp probe obtained by digestion of
CC Enterococcus faecalis genomic DNA with the restriction enzyme HindIII.
CC The probes (AAQ55133-48 and AAQ55752-58) represent other probes derived,
CC by HindIII digestion of the genomes of S.aureus, S.epidermis,

```
CC   Enterococcus faecalis, Pseudomonas aeruginosa, E.coli, Klebsiella
CC   pneumoniae or Enterobacter cloacae. The probes can be used to detect
CC   their respective microorganisms in clinical samples. (Updated on 25-MAR-
CC   2003 to correct PN field.)
XX
SQ   Sequence 3480 BP; 1068 A; 569 C; 698 G; 1145 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 49.6;  DB 2;  Length 3480;
  Best Local Similarity   64.8%;  Pred. No. 0.63;
  Matches   92;  Conservative    0;  Mismatches   44;  Indels    6;  Gaps    1;

Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
              | || ||| ||||||||  | || ||||||  || || |||||||  |  ||    ||  |||
Db       2797 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 2856

Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
              ||||      ||    | || | | |  |||   |||  ||  ||| ||||||  |||  |
Db       2857 GTAT------TCTTTTACAAGTCAGTGAAGAAGATGAACTAATTATTTCAGATGATGGTT 2910

Qy        125 CTACTGATGGTACAGAGGAAAT 146
              ||||||||  |||   ||||||
Db       2911 CTACTGATCATACGTTGGAAAT 2932


RESULT 10
ABZ77359
ID   ABZ77359 standard; DNA; 3480 BP.
XX
AC   ABZ77359;
XX
DT   28-MAY-2003  (first entry)
XX
DE   Nucleotide sequence designated EF-7.
XX
KW   Phagocyte; microorganism; immune function; infection; bacteria;
KW   septicemia; fungi; ss.
XX
OS   Enterococcus faecalis.
XX
PN   WO2002101037-A1.
XX
PD   19-DEC-2002.
XX
PF   27-MAY-2002; 2002WO-JP005106.
XX
PR   31-MAY-2001; 2001JP-00165954.
XX
PA   (FUSO ) FUSO PHARM IND LTD.
XX
PI   Ohno T,  Matsuhisa A,  Keshi H,  Abe K,  Sugimoto N,  Ueyama H;
PI   Eda S,  Uehara H,  Iwami T,  Yamamoto S,  Araki H;
XX
DR   WPI; 2003-148802/14.
XX
PT   Preparing phagocytes which have taken up a foreign microorganism, useful
PT   for comprises contacting the phagocytes with the microorganism then
PT   isolating the phagocytes.
XX
PS   Example 7; Page 130-132; 155pp; Japanese.
XX
CC   The specification describes a method of preparing phagocytes which have
CC   taken up a foreign microorganism. the method comprises contacting the
CC   phagocytes with the microorganism before isolating the phagocytes. The
CC   method is useful for in vitro evaluation of function of phagocytes or
CC   microbicidal ability, and also for evaluating immune function and
CC   differentiation efficiency of phagocytes, screening modulators of
CC   phagocytic function, in management of clinical examination and as
CC   indication in infection diagnosis including infection state by foreign
CC   microorganisms like bacteria causing septicemia and by fungi. The present
CC   sequence is used in the course of the invention
XX
SQ   Sequence 3480 BP; 1068 A; 569 C; 698 G; 1145 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 49.6;  DB 8;  Length 3480;
  Best Local Similarity   64.8%;  Pred. No. 0.63;
  Matches   92;  Conservative    0;  Mismatches   44;  Indels    6;  Gaps    1;

Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
              | || ||| ||||||||  | || ||||||  || || |||||||  |  ||    ||  |||
Db       2797 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 2856

Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
              ||||      ||    | || | | |  |||   |||  ||  ||| ||||||  |||  |
Db       2857 GTAT------TCTTTTACAAGTCAGTGAAGAAGATGAACTAATTATTTCAGATGATGGTT 2910

Qy        125 CTACTGATGGTACAGAGGAAAT 146
              ||||||||  |||   ||||||
Db       2911 CTACTGATCATACGTTGGAAAT 2932


RESULT 11
AAL51847
ID   AAL51847 standard; DNA; 3480 BP.
XX
AC   AAL51847;
XX
DT   01-MAY-2003  (first entry)
XX
DE   Enterococcus faecalis microbe identification-related DNA sequence #3.
XX
KW   Microbe identification; phagocyte isolation; membrane permeability; ds;
KW   infection; blood poisoning; septicaemia.
XX
OS   Enterococcus faecalis.
XX
PN   WO200299133-A1.
XX
PD   12-DEC-2002.
XX
PF   27-MAY-2002; 2002WO-JP005107.
XX
PR   31-MAY-2001; 2001JP-00165929.
XX
PA   (FUSO ) FUSO PHARM IND LTD.
PA   (OHNO/) OHNO T.
XX
PI   Ohno T,  Matsuhisa A;
XX
DR   WPI; 2003-148679/14.
XX
PT   Detection of specific microbial DNA from treated phagocytes comprises
PT   probe hybridisation for identification of infection-causing microbes in
PT   clinical samples.
XX
PS   Example 7; Page 105-107; 129pp; Japanese.
XX
CC   The invention comprises a method for identifying microbes causing
CC   infection in clinical specimens. The method involves: isolating and
CC   immobilising phagocytes from the clinical specimens; the phagocytes are
CC   then treated to enhance their membrane permeability with lysostaphin,
CC   lysozyme, N-acetylmuramidase or zymolase; the exposed DNA is then
CC   hybridised with a DNA probe. The method of the invention is useful for
CC   the identification of the microbial type causing infection and blood
CC   poisoning (e.g. septicaemia). The present DNA sequence represents a
CC   microbial nucleotide sequence used in an example of the invention
XX
SQ   Sequence 3480 BP; 1068 A; 569 C; 698 G; 1145 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 49.6;  DB 10;  Length 3480;
  Best Local Similarity   64.8%;  Pred. No. 0.63;
  Matches   92;  Conservative    0;  Mismatches   44;  Indels    6;  Gaps    1;

Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
              | || ||| ||||||||  | || ||||||  || || |||||||  |  ||    ||  |||
Db       2797 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 2856
```

```
Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
              ||||      ||    | || | | |  |||   |||  ||  ||| |||||| |||  |
Db       2857 GTAT------TCTTTTACAAGTCAGTGAAGAAGATGAACTAATTATTTCAGATGATGGTT 2910

Qy        125 CTACTGATGGTACAGAGGAAAT 146
              ||||||||  |||   ||||||
Db       2911 CTACTGATCATACGTTGGAAAT 2932


RESULT 12
AAT58840_0
WP Sequence split into 6 fragments  LOCUS AAT58840 Accession Aat58840
WP     Fragment Name      Begin     End
WP     AAT58840_0             1     110000
WP     AAT58840_1        100001     210000
WP     AAT58840_2        200001     310000
WP     AAT58840_3        300001     410000
WP     AAT58840_4        400001     510000
WP     AAT58840_5        500001     580073
ID   AAT58840 standard; DNA; 580073 BP.
XX
AC   AAT58840;
XX
DT   27-MAR-1997  (first entry)
XX
DE   Mycoplasma genitalium genome.
XX
KW   M. genitalium; DNAA; DNA gyrase; origin of replication;
KW   megabase shotgun sequencing method; open reading frame; ORF; ss.
XX
OS   Mycoplasma genitalium.
XX
FH   Key             Location/Qualifiers
FT   CDS             1026. .1829
FT                   /*tag=  gk
FT                   /label= MG001
FT                   /note= "Previously identified as MORF-20072, the encoded
FT                   protein shows 100 percentage identity to DNA polymerase
FT                   III beta subunit (dnaN) from M. genitalium"
FT   CDS             1829. .2761
FT                   /*tag=  gb
FT                   /label= MG002
FT                   /note= "Previously identified as MORF-19817, the encoded
FT                   protein shows 61.67 percentage identity to heat shock
FT                   protein (dnaJ) from Lactococcus lactis"
FT   CDS             2846. .4798
FT                   /*tag=  gc
FT                   /label= MG003
FT                   /note= "Previously identified as MORF-19818 and MORF-
FT                   20073, the encoded protein shows 99.38 percentage
FT                   identity to DNA gyrase subunit B (gyrB) from M.
FT                   genitalium"
FT   CDS             4813. .7323
FT                   /*tag=  gd
FT                   /label= MG004
FT                   /note= "Previously identified as MORF-19819 and MORF-
FT                   20074, the encoded protein shows 99.88 percentage
FT                   identity to DNA gyrase subunit A (gyrA) from M.
FT                   genitalium"
FT   CDS             7295. .8548
FT                   /*tag=  gl
FT                   /label= MG005
FT                   /note= "Previously identified as MORF-20075, the encoded
FT                   protein shows 66.34 percentage identity to seryl-tRNA
FT                   synthetase (serS) from B. subtilis"
FT   CDS             8552. .9184
FT                   /*tag=  a
FT                   /label= MG006
FT                   /note= "Previously identified as MORF-20076, the encoded
FT                   protein shows 27.59 percentage identity to thymidylate
FT                   kinase (CDC8) from Saccharomyces cerevisiae"
FT   CDS             9157. .9921
FT                   /*tag=  gm
FT                   /label= MG007
FT                   /note= "Previously identified as MORF-19820, the encoded
FT                   protein shows 45.39 percentage identity to DNA polymerase
FT                   III subunit (dnaH) from B. subtilis"
FT   CDS             9924. .11252
FT                   /*tag=  gn
FT                   /label= MG008
FT                   /note= "Previously identified as MORF-20077, the encoded
FT                   protein shows 59.78 percentage identity to thiopene and
FT                   furan oxidiser (tdhF) from B. subtilis"
FT   CDS             11252. .12040
FT                   /*tag=  b
FT                   /label= MG009
FT                   /note= "Previously identified as MORF-20078, the encoded
FT                   protein shows 35.43 percentage identity to the Bacillus
FT                   subtilis hypothetical protein covered in accession number
FT                   GB:D26185_102"
FT   CDS             12069. .12725
FT                   /*tag=  c
FT                   /label= MG010
FT                   /note= "Previously identified as MORF-20079, the encoded
FT                   protein shows 25.73 percentage identity to DNA primase
FT                   (dnaE) from Clostridium acetobutylicum"
FT   CDS             complement(12702. .13565)
FT                   /*tag=  go
FT                   /label= MG011
FT                   /note= "Previously identified as MORF-19821 and MORF-
FT                   19822"
FT   CDS             complement(13570. .14247)
FT                   /*tag=  d
FT                   /label= MG012
FT                   /note= "Previously identified as MORF-20080, the encoded
FT                   protein shows 31.50 percentage identity to the ribosomal
FT                   protein S6 modification protein (rimK) from Escherichia
FT                   coli"
FT   CDS             complement(14396. .15217)
FT                   /*tag=  e
FT                   /label= MG013
FT                   /note= "Previously identified as MORF-19823, MORF-20080
FT                   and MORF-20081, the encoded protein shows 33.04
FT                   percentage identity to 5,10-methylene-tetra- hydrofolate
FT                   dehydrogenase (folD) from E. coli"
FT   CDS             15556. .17427
FT                   /*tag=  gp
FT                   /label= MG014
FT                   /note= "Previously identified as MORF-20082, the encoded
FT                   protein shows 52.63 percentage identity to transport ATP-
FT                   binding protein (msbA) from E. coli"
FT   CDS             17474. .19243
FT                   /*tag=  f
FT                   /label= MG015
FT                   /note= "Previously identified as MORF-20084, the encoded
FT                   protein shows 32.23 percentage identity to transport ATP-
FT                   binding protein (msbA) from E. coli"
FT   CDS             19253. .19759
FT                   /*tag=  pc
FT                   /label= MG016
FT                   /note= "Previously identified as MORF-19824"
FT   CDS             19761. .198165
FT                   /*tag=  bs
FT                   /label= MG166
FT                   /note= "Previously identified as MORF-19901 and MORF-
FT                   20206, the encoded protein shows 46.99 percentage
FT                   identity to ribosomal protein L6 (rpL6) from M.
FT                   capricolum"
FT   CDS             19825. .20355
FT                   /*tag=  pd
FT                   /label= MG017
FT                   /note= "Previously identified as MORF-19825"
FT   CDS             21063. .22346
FT                   /*tag=  gq
FT                   /label= MG018
FT                   /note= "Previously identified as MORF-20085, the encoded
```

```
FT                   protein shows 30.09 percentage identity to helicase
FT                   (mot1) from S. cerevisiae"
FT   CDS             22388. .23557
FT                   /*tag= gr
FT                   /label= MG019
FT                   /note= "Previously identified as MORF-20086, the encoded
FT                   protein shows 51.11 percentage identity to heat shock
FT                   protein (dnaJ) from L. lactis"
FT   CDS             23541. .24467
FT                   /*tag= gs
FT                   /label= MG020
FT                   /note= "Previously identified as MORF-20088, the encoded
FT                   protein shows 55.79 percentage identity to proline
FT                   iminopeptidase (pip) from Neisseria gonorrheae"
FT   CDS             24467. .26005
FT                   /*tag= gt
FT                   /label= MG021
FT                   /note= "Previously identified as MORF-20089, the encoded
FT                   protein shows 58.89 percentage identity to methionyl-tRNA
FT                   synthetase (metS) from B. subtilis"
FT   CDS             26035. .26472
FT                   /*tag= gu
FT                   /label= MG022
FT                   /note= "Previously identified as MORF-20091, the encoded
FT                   protein shows 49.26 percentage identity to RNA polymerase
FT                   delta subunit (rpoE) from B. subtilis"
FT   CDS             26478. .27344
FT                   /*tag= g
FT                   /label= MG023
FT                   /note= "Previously identified as MORF-20092, the encoded
FT                   protein shows 45.96 percentage identity to fructose-
FT                   bisphosphate aldolase (tsr) from B. subtilis"
FT   CDS             27345. .28448
FT                   /*tag= h
FT                   /label= MG024
FT                   /note= "Previously identified as MORF-19826 and MORF-
FT                   20093, the encoded protein shows 46.84 percentage
FT                   identity to GTP-binding protein from E. coli"
FT   CDS             28651. .29547
FT                   /*tag= gv
FT                   /label= MG025
FT                   /note= "Previously identified as MORF-20094, the encoded
FT                   protein shows 54.08 percentage identity to TrsB from
FT                   Yersinia enterocolitica"
FT   CDS             29551. .30123
FT                   /*tag= gw
FT                   /label= MG026
FT                   /note= "Previously identified as MORF-20095, the encoded
FT                   protein shows 47.24 percentage identity to elongation
FT                   factor P (efp) from E. coli"
FT   CDS             30092. .30547
FT                   /*tag= pe
FT                   /label= MG027
FT                   /note= "Previously identified as MORF-19827"
FT   CDS             30547. .31152
FT                   /*tag= pf
FT                   /label= MG028
FT                   /note= "Previously identified as MORF-19828"
FT   CDS             complement(31142. .31702)
FT                   /*tag= gx
FT                   /label= MG029
FT                   /note= "Previously identified as MORF-19829, the encoded
FT                   protein shows 45.05 percentage identity to the S. aureus
FT                   hypothetical protein given in GB:L19300_1"
FT   CDS             complement(31704. .32324)

  Query Match             5.0%;  Score 49.6;  DB 2;  Length 110000;
  Best Local Similarity   46.0%;  Pred. No. 1.1;
  Matches  209;  Conservative    0;  Mismatches  239;  Indels    6;  Gaps    1;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||     ||||  |   ||| || ||||| ||| |   |  || | |||||  |     |
Db      67651 ATGAAATTATCTGTAATTATACCTACTTACAATTGTGCATCATTTATTGAAAAAGCAATT 67710

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              |||   ||   |||  |   |   | |||| || || ||  | |   ||    ||| ||
Db      67711 AATTCAATTGTTAAAAATAGACCTAATGATTTGGAAATAGAAGTTTTAATTATTGATGAT 67770

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              |  || | |||   |||  |  || | ||| |                || |  ||||     |
Db      67771 GGATCAATTGACAATACTAACAAAGTTATTAAGAAAATTCAAGACCAAATTAATAATTTA 67830

Qy        181 ATAATATATCGGCGT------AATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCT 234
              |   | |       |       | ||  | ||||   ||    | ||  |||| | ||
Db      67831 ACTTTGCAGTATTTTTACAAAAGTAATGGTAACTGGGGTAGTGTTATTAATTATGTTAGA 67890

Qy        235 TCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCG 294
                || || || ||||| |          |||| |  |||||||   | || |
Db      67891 AACAATAAACTAGCAAAAGGGGAATGAGTAACAGTATTGGATAGTGATGACATTTTTTCA 67950

Qy        295 AAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGT 354
              ||| |  |  |  | || || ||||  |||  | |    |||  | ||       |  |
Db      67951 AAAAAAACAATTTCTATTTTTCAAAAATATGCCCAAAAACAAAGATATGATGCGATTATT 68010

Qy        355 GACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTC 414
                  | | |    | ||  |      ||  |  | ||  |||      | |  || |
Db      68011 TTTGACTACTATAAATGCTGAAAAAAGTTTTTGTGAAAAATTCCTACCTATGCAAGGTTT 68070

Qy        415 AGAACAGATGATGAACTTTATGTGTTTAATAATA 448
              |||| |||   | ||  | |  ||   ||  | |
Db      68071 AGAAAAGAAATTAAAGGTGAATTGAAAAAACAAA 68104


RESULT 13
AAX13380/c
ID   AAX13380 standard; DNA; 2012 BP.
XX
AC   AAX13380;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:443.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US008985.
XX
PR   06-MAY-1997;   97US-0044031P.
PR   16-MAY-1997;   97US-0046655P.
PR   14-NOV-1997;   97US-0066009P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Kunsch CA,  Dillon PJ,  Barash SC;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides -
PT   used to develop products for the detection of Enterococcus and for use in
PT   vaccines for prevention or attenuation of Enterococcus infection.
XX
PS   Claim 1; Page 1673-1675; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence of
```

```
CC   Enterococcus faecalis in samples. They can also be used for diagnosing
CC   Enterococcal infection in an animal and monitoring progression of
CC   disease, and for identifying agents which can be used to modulate the
CC   growth or pathogenicity of Enterococcus faecalis, or another related
CC   organism, in vivo or in vitro. In particular the polypeptides encoded by
CC   the Enterococcus faecalis nucleotide sequences can be used in vaccines to
CC   prevent or attenuate an Enterococcal infection
XX
SQ   Sequence 2012 BP; 578 A; 423 C; 320 G; 675 T; 0 U; 16 Other;

  Query Match             4.9%;  Score 49.2;  DB 2;  Length 2012;
  Best Local Similarity   64.1%;  Pred. No. 0.7;
  Matches   91;  Conservative    1;  Mismatches   44;  Indels    6;  Gaps    1;

Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
              | || ||| |||||||| | || |||||| || || ||||||| |  ||     ||  |||
Db       1890 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 1831

Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
              ||||   |   ||:|      | | |  |||   |||  ||  ||| ||||||| |||  |
Db       1830 GTATTCTTTTACARG------TCAGTGAAGARGATGAACTAATTATTTCAGATGATGGTT 1777

Qy        125 CTACTGATGGTACAGAGGAAAT 146
              ||||||||  |||   ||||||
Db       1776 CTACTGATCATACGTTGGAAAT 1755


RESULT 14
ABS99175/c
ID   ABS99175 standard; DNA; 2012 BP.
XX
AC   ABS99175;
XX
DT   18-DEC-2002  (first entry)
XX
DE   Enterococcus faecalis contig sequence #443.
XX
KW   Computer readable medium; Enterococcus faecalis; microbe; growth;
KW   pathogenicity; vaccine; resistance; Enterococcal infection; commercial;
KW   therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;
KW   biochip technology; antibacterial; modulator of nucleic acid expression;
KW   contig; ds.
XX
OS   Enterococcus faecalis.
XX
PN   US2002120116-A1.
XX
PD   29-AUG-2002.
XX
PF   04-MAY-1998;   98US-00070927.
XX
PR   04-MAY-1998;   98US-00070927.
XX
PA   (KUNS/) KUNSCH C A.
PA   (DILL/) DILLON P J.
PA   (BARA/) BARASH S.
XX
PI   Kunsch CA,  Dillon PJ,  Barash S;
XX
DR   WPI; 2002-750065/81.
XX
PT   Computer readable medium having recorded on it a Enterococcus faecalis
PT   nucleotide sequence useful for detecting diseases related to Enterococcus
PT   infections in animals.
XX
PS   Claim 1; Page; 119pp; English.
XX
CC   The present invention relates to a new computer readable medium with an
CC   Enterococcus faecalis nucleotide sequence. The invention is useful to
CC   diagnose the presence of E.faecalis in a sample or determining the
CC   presence of a specific microbe in a sample. The invention is also useful
CC   for modulating the growth or pathogenicity of E.faecalis, in a vaccine to
CC   confer resistance to Enterococcal infection, for commercial, therapeutic
CC   and industrial purposes, and for fermenting a particular sugar source or
CC   to produce a particular metabolite. The invention is useful for detecting
CC   diseases related to Enterococcus infections in animals, and for detecting
CC   E.faecalis using biochip technology. The present nucleic acid sequence
CC   represents an Enterococcus faecalis contig DNA sequence of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at http.sequdata.uspto.gov
XX
SQ   Sequence 2012 BP; 578 A; 423 C; 320 G; 675 T; 0 U; 16 Other;

  Query Match             4.9%;  Score 49.2;  DB 6;  Length 2012;
  Best Local Similarity   64.1%;  Pred. No. 0.7;
  Matches   91;  Conservative    1;  Mismatches   44;  Indels    6;  Gaps    1;

Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
              | || ||| |||||||| | || |||||| || || ||||||| |  ||     ||  |||
Db       1890 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 1831

Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
              ||||   |   ||:|      | | |  |||   |||  ||  ||| ||||||| |||  |
Db       1830 GTATTCTTTTACARG------TCAGTGAAGARGATGAACTAATTATTTCAGATGATGGTT 1777

Qy        125 CTACTGATGGTACAGAGGAAAT 146
              ||||||||  |||   ||||||
Db       1776 CTACTGATCATACGTTGGAAAT 1755


RESULT 15
ADH82927
ID   ADH82927 standard; DNA; 477 BP.
XX
AC   ADH82927;
XX
DT   22-APR-2004  (first entry)
XX
DE   Enterococcus faecalis polynucleotide #812.
XX
KW   Enterococcus faecalis infection; transcription regulatory element;
KW   antibacterial; gene; ds.
XX
OS   Enterococcus faecalis.
XX
PN   US6617156-B1.
XX
PD   09-SEP-2003.
XX
PF   13-AUG-1998;   98US-00134000.
XX
PR   15-AUG-1997;   97US-0055778P.
XX
PA   (DOUC/) DOUCETTE-STAMM L A.
PA   (BUSH/) BUSH D.
XX
PI   Doucette-Stamm LA,  Bush D;
XX
DR   WPI; 2003-895394/82.
DR   P-PSDB; ADH86332.
XX
PT   New nucleic acid comprising a sequence encoding an Enterococcus fecalis
PT   polypeptide, useful for preparing a composition for diagnosing or
PT   treating E. fecalis infection.
XX
PS   Disclosure; SEQ ID NO 812; 193pp; English.
XX
CC   The invention relates to Enterococcus faecalis polynucleotides and
CC   polypeptides. The invention also relates to a recombinant expression
CC   vector comprising a polynucleotide operably linked to a transcription
CC   regulatory element, a cell comprising a recombinant vector, a method for
CC   producing an E. faecalis polypeptide, an isolated nucleic acid comprising
CC   a sequence not given in the specification, a recombinant vector
CC   comprising the nucleic acid and a cell comprising the recombinant vector.
CC   The polynucleotides can be used to detect the presence of E. faecalis in
```

CC a sample. The sequences are useful for preparing a composition for
CC diagnosing or treating Enterococcus faecalis infection. This sequence
CC represents an E. faecalis polynucleotide of the invention.
XX
SQ Sequence 477 BP; 155 A; 72 C; 104 G; 146 T; 0 U; 0 Other;

Query Match 4.8%; Score 48; DB 10; Length 477;
Best Local Similarity 64.1%; Pred. No. 1;
Matches 91; Conservative 0; Mismatches 45; Indels 6; Gaps 1;

```
Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
              | || ||| |||||||| | || |||||| || || ||||||| |  ||    ||  |||
Db          5 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 64

Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
              ||||       ||    | || | | |  |||   |||  ||  ||| |||||| |||  |
Db         65 GTAT------TCTTTTACAAGTCAGTGAAGAAGATGAACTAATTATTTCAGATGATGGTT 118

Qy        125 CTACTGATGGTACAGAGGAAAT 146
               |||||||  |||   ||||||
Db        119 ATACTGATCATACGTTGGAAAT 140
```

Search completed: May 11, 2006, 09:43:04
Job time : 460.564 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 09:43:19 ; Search time 3121.9 Seconds
(without alignments)
15016.694 Million cell updates/sec

Title: US-09-423-093B-4_COPY_14059_15060
Perfect score: 1002
Sequence: 1 atgcttatatcattttgtat.....ctaaacggataattaagtag 1002

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	62.6	6.2	1101	10	CNS0039G	AL063921 Drosophil
c	2	58.4	5.8	1101	10	CNS00EO7	AL069440 Drosophil
	3	57.6	5.7	1101	10	CNS00EVL	AL069706 Drosophil
	4	57.4	5.7	1101	10	CNS0006J	AL062049 Drosophil
c	5	56.8	5.7	1101	10	CNS00EVL	AL069706 Drosophil
c	6	55.2	5.5	1043	10	CNS0145P	AL103735 Drosophil
	7	55	5.5	1101	10	CNS000D1	AL065414 Drosophil
	8	54.8	5.5	1225	10	CNS0161D	AL106171 Drosophil
	9	54.4	5.4	922	10	CNS015W0	AL105978 Drosophil
c	10	54.4	5.4	1101	10	CNS017ZQ	AL108704 Drosophil
	11	54.2	5.4	1101	10	CNS0106X	AL098595 Drosophil
c	12	53.4	5.3	562	9	AZ522603	AZ522603 208PbF09
	13	53	5.3	1101	10	CNS0182P	AL108811 Drosophil
c	14	52.8	5.3	1101	10	CNS001FB	AL060732 Drosophil
	15	52.6	5.2	1201	10	CNS0162X	AL106227 Drosophil
	16	52.6	5.2	1224	8	DN809265	DN809265 76946254
	17	51	5.1	1084	11	CNS071NH	AL425139 clone BA0
c	18	51	5.1	1203	10	CNS015WU	AL106008 Drosophil
	19	50.8	5.1	781	10	CNS009DO	AL053444 Drosophil
	20	50.8	5.1	907	9	BH154707	BH154707 ENTRL79TR
	21	50.6	5.0	969	10	CL077158	CL077158 CH216-143
	22	50.4	5.0	1031	10	CNS00CF2	AL059199 Drosophil
	23	50.4	5.0	1101	10	CNS0039G	AL063921 Drosophil
	24	50.4	5.0	1200	10	CNS016CO	AL106578 Drosophil
c	25	50.2	5.0	1225	8	DN734458	DN734458 CNB75-G09
c	26	50	5.0	1101	10	CNS00FYG	AL071206 Drosophil
	27	49.8	5.0	879	9	AZ545134	AZ545134 ENTDL16TF
	28	49.8	5.0	883	9	BH165994	BH165994 ENTTN78TR
	29	49.8	5.0	1099	10	CNS00DW8	AL076311 Drosophil
	30	49.8	5.0	1277	9	CC253231	CC253231 CH261-180
c	31	49.8	5.0	4115	4	CR859173	CR859173 Pongo pyg
	32	49.4	4.9	879	9	AZ672806	AZ672806 ENTGA19TR
c	33	49.4	4.9	928	10	CNS00DKY	AL071865 Drosophil
c	34	49.4	4.9	1101	10	CNS016LI	AL106896 Drosophil
c	35	49.4	4.9	1103	1	AJ926366	AJ926366 AJ926366
c	36	49.4	4.9	1204	10	CNS016E2	AL106628 Drosophil
c	37	49.2	4.9	821	10	CZ970815	CZ970815 170475 To
	38	49.2	4.9	842	9	AZ681842	AZ681842 ENTIA03TF
	39	49.2	4.9	857	9	AZ673200	AZ673200 ENTIA73TF
c	40	49.2	4.9	1542	10	AG386981	AG386981 Mus muscu
	41	49	4.9	979	10	CNS0161W	AL106190 Drosophil
c	42	49	4.9	1101	10	CNS000B8	AL063632 Drosophil
c	43	49	4.9	1101	10	CNS0039L	AL063926 Drosophil
	44	49	4.9	1173	1	AJ926134	AJ926134 AJ926134
c	45	48.6	4.9	604	7	CO312735	CO312735 EK229834.

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

```
  Query Match             6.2%;  Score 62.6;  DB 10;  Length 1101;
  Best Local Similarity   17.9%;  Pred. No. 0.0056;
  Matches  118;  Conservative  271;  Mismatches  265;  Indels    5;  Gaps    1;

Qy        250 AATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGACTCGTTAGCG 309
              :::|:: :: :| ::: ::| ::|   :: :: :: |    :  :: ||  :  |::
Db       1098 RRWGDDTWDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWW 1039

Qy        310 ATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGAAAAGAGACC 369
              ::: |:  :: ::::    |:| | :   : |:: ::|  ::   :: ::|| : :||
Db       1038 WWWATWDTWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGR 979

Qy        370 GGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACAGATGATGAA 429
              ::| ::: :::::::|:::::  |       : |:  |: ::: :  ::  :||:::: ::
Db        978 RDGGRKRKDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADD 919

Qy        430 CTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATCTATT 489
               :   ::::: :  :|::: : | :::::::::::: :|: |  :   |:  |::: : :
Db        918 DDGAGDKDDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWA 859

Qy        490 GGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGGGATGCCATT 549
              : :  | :|  ::   :: ::|  : :::|: :| : |:::: :   ::|||| :  : :
Db        858 DADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRK 799

Qy        550 GATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGCGTA 609
              :|::: :| : ::    : |:     ||  |::  :  ::|:|:|:: |:: :::: :::
Db        798 RADDKRDAADDRDDAATWTTW-----TTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDD 744

Qy        610 TTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATATGCCGAGGA 669
               : | ||  :  ::  ::  :  :  :::|:|::| :::  : : | :|   :  |::::
Db        743 RDRAGTAGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRR 684

Qy        670 GATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCA 729
              ::::|| :  : :: :| ::   : : :: :  ::: : |    : | :| ::::|:: :
Db        683 RRDRATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAW 624

Qy        730 TATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAA 789
              :: ::::    :: :::  :|:|: : |: ::|  | |::   : |:||  :: :||:|
Db        623 DWAKWDWKTRADRWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKT 564

Qy        790 AATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCATGTTATGGC 849
                 :||:   ||:: : | |      || : |      |||    | :   |  ::|
Db        563 TTATWTTWAARAAWWAWWAWATTTATWTTTWTTTWTTTTTTTTTTTWTTTAAWWWAAWWT 504

Qy        850 AATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAA 908
              | :| : |  :|||| | | || | |   |||:|:  ||:   :  || |: |  ::|
Db        503 ATWAAWTAAAWAAAAAAAWAATTTTTTTTTTTTTWAWTAAAWWTAWTWTTTTWTTTTWWA 445


RESULT 2
CNS00EO7/c
LOCUS       CNS00EO7                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069440
VERSION     AL069440.1  GI:4949583
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29P01"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.8%;  Score 58.4;  DB 10;  Length 1101;
  Best Local Similarity   37.9%;  Pred. No. 0.045;
  Matches  192;  Conservative   64;  Mismatches  246;  Indels    5;  Gaps    2;

Qy        428 AACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATCTA 487
              |:: :|:| |  :|| |  |  |||: |||   :|   | |:     | :   || : ||
Db       1019 AMMAYTWTTTTAWTATTTTTTTTTTTMAATATYCWATATTTTMAATACAWWAYATTTWWTA 960

Qy        488 TTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGGGATGCCA 547
              |   |  | : |||   |||  ||    |||:| | ||||||:  |   |    |:   :
Db        959 TACATAATWWTTTTTTATATACAA----TTTWAAAATAAAAAWTAACWWAATTTAWAAAW 904

Qy        548 TTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGCG 607
                  :|||       |  | |: |  || |:   ::    |    :|||  :| |  |
Db        903 CATTWTTTTTAAWWTAATTTWATTACAWTWTTAWWTAAAAAAATWTTAAAWTAAAWAAAA 844

Qy        608 TATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATATGCCGAG 667
               :: | |:|    |   :   |:|   ::||||:|| |  :  | |  ||  :
Db        843 ARWATTAWAATTTAYATWATATWTAAAWWTATAWATTATTMWWAATWTTATAWNATTTTT 784

Qy        668 GAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTG 727
              :||:  :|: |:   |  |      |:    | |  ||    | ||  |||||||
Db        783 TTTWTAWAAWTWTTWATWAWTAATTTTAAWTWWTTAAATAAAAAAAATTTATTTTTATTT 724

Qy        728 CATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTG 787
              ||||:|||:||   |  | | ||::     | |: || |: :||||:  :  |:||
Db        723 WTTATTWAAAWTTTWTTTTWAATTWWYTTTAATAWTTAAAWTWTTAAAWAWWTAAWTTA 664

Qy        788 AAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCATGTTATG 847
              || |  |:|   ||||| : : |  ||   | ||:| :   |    | |   :: :| :
Db        663 AATAAATWTATTTAAAATWTWAAWTATAAATTTAWADWTTATAWWTTTTTTTWWCWTTWA 604

Qy        848 GCAATAGTGATGAAAAAAGAGATTTATCTGA-ATTTTATGCAAAGCTAGGTTGTAATAAA 906
              |     || |||||  | |||: | | | ||:::||  |||  |:  || |    ||
Db        603 ATATATAAAATWAAAAATTAAATTWTTTTTATATWWWATAAAAAMATWWWTTATTTATAA 544

Qy        907 AATATGATCAACACTGTACTTCGATTT 933
              :: || ||  | | |  | ||  ||||
Db        543 WWAATTATTTAWAWTTGATTTTTATTT 517


RESULT 3
CNS00EVL
LOCUS       CNS00EVL                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
```

```
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             5.7%;  Score 57.6;  DB 10;  Length 1101;
  Best Local Similarity   35.8%;  Pred. No. 0.067;
  Matches  126;  Conservative   56;  Mismatches  170;  Indels    0;  Gaps    0;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              :|  ||:::| |||   || |  :|  ||:|| | |||| || |   | || | |: |
Db        661 WTWATTWWWTAATTAAATAATTTWAAATAWAAAAAAAAAAAAAAAAWATWAAWAATWATAW 720

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              |  | |  |||:||| |  :||:| |:|:: ||    ::         | :   : ||::
Db        721 ATAAWTTAAAAWAATAAAAWAAWAATWAWWATAATATWWATATATAWTTWTAWWWATWWA 780

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
                | ||   |:  || | :  |:|| | | :   :   |  :| |: ||::||:    |
Db        781 WWWTATAWATAWAATAWAAWAWAWATAAATAWATAWATWAAAWAWAWATAWWATWATATA 840

Qy        181 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 240
              ::||||:|:       :||   |:||   |     | |    |:||:: :|  |: | :|
Db        841 WWAATAWAWAAAAAATWTAATATWAATWATAAWAAAAAATAWAWTTWWTWTTTTWWAWWA 900

Qy        241 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 300
              |        |       :  |       | ||     : | ||  |  | ||   : : |
Db        901 TATAAAWAWATAWAAAWAAAAAAAAAAAAATAAWAWWWTWATATTWTTATTAAAWTWTAT 960

Qy        301 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTAT 352
              :  ||:   ::|:|: | | ||:| |  ||    ::  : |:||||||||||
Db        961 WWATTWATTWWAWTWTATAWTTWTATWTATATATWWTAWTAWATATATTTAT 1012


RESULT 4
CNS0006J
LOCUS       CNS0006J               1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR01M22 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL062049
VERSION     AL062049.1  GI:4938511
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR01M22"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.7%;  Score 57.4;  DB 10;  Length 1101;
  Best Local Similarity   22.6%;  Pred. No. 0.074;
  Matches   78;  Conservative  135;  Mismatches  131;  Indels    1;  Gaps    1;

Qy        657 AATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAAT 716
              |: : |  | : ::| ||:    ::|: ::::|  :: ::|: |  | |  :|  : | :
Db        718 ARARGGRGGGRAWRAAAAKTKRKKKTKTKRKRAAGRRARRAWAGAAAAAAARAAADAATK 777

Qy        717 TGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAA 776
              || |: :| :: | ::::: | : |  :|| :  ::|||  |::   :  | ||  : | |
Db        778 TGKTWAKAAWRTAAWKKDWKATDAAAAKAAARTRDWTTAKAKDKTKWGAAAATGAWGAGA 837

Qy        777 ACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTAT 836
              | |::   : ::  : :      : ::|:| : :  |:  ||| ::   | :: :   ::
Db        838 ARGRRWGRKGKDGTRARRARGAGDGWDAKAAWAWAAAWAATTARWDTATAADDRRAKAWD 897

Qy        837 GGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGG 896
              :: |   :::::  : ||:| ||::::|: : :: :|:  ||:   :|:: |  : : :|
Db        898 KRAAAAADRDKDKGRRAGWG-TGWRRRARARTRKAKAKRRGARAGAKAKRRAKGDAWKDG 956

Qy        897 TTGTAATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAA 956
              :::::::::||:|   : | ::   ::| |  :|  ::::::|  :: :::|: : : ::
Db        957 WKKKWDKDAAWAAKAKGTGDWKKRARKAKTGWKARADKDRWRAAADDTRWRTKAWWWTDW 1016

Qy        957 AAATATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTA 1001
              ::||||::   ::  :  :|::|||: ::||  :|:::| | :||
Db       1017 RWATATKDKDTKDKAKKGKAWKTTTRTDRAAKAKAKDDTAATRTA 1061


RESULT 5
CNS00EVL/c
LOCUS       CNS00EVL               1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
```

```
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:7227"
                      /clone="BACR29B23"
                      /clone_lib="RPCI-98"
                      /note="end : T7"
ORIGIN

  Query Match             5.7%;  Score 56.8;  DB 10;  Length 1101;
  Best Local Similarity   33.0%;  Pred. No. 0.099;
  Matches  169;  Conservative   87;  Mismatches  254;  Indels    2;  Gaps    1;

Qy        442 AATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATCTATTGGTGGTGTATTT 501
              |:||  :|:|:: ||:   ::|| :|: | | || :   |:  ::|:     | | |||
Db        981 AWTATAWAWTWWAATWAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTT 922

Qy        502 AGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGGGATGCCATTGATTTTGATGCG 561
                |:| |:  |: :|| ||| :: ::||| :|  |          || | |:  |
Db        921 TTTTWTTTWTATWTWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWAT 862

Qy        562 TCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGCGTATTTAATACGCCA 621
                :  ||   : :|  |::  |  : |::|  : :|  | :   |||:|| |     :
Db        861 ATTAWATTTTTTWTWTATTWWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTW 802

Qy        622 GGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATATGCCGAGGAGATAATGATAGT 681
              |   ||:|  :|   :::| :| :    :  : ||||    |    ||::| ||  |
Db        801 TWTTWTATTWTATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWT 742

Qy        682 TTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTA 741
              |:    |  :    ||:    : |  | |:||:||:  |||| ||   | |:||   : |
Db        741 TWTTTTATTWTTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTTWTATTTWAA 682

Qy        742 GCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTA 801
                || | | || :: | |:|:    |   |  : ||  |   | | ||||   |||  |:
Db        681 ATTATTTAATTAWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAA--ATTTTATW 624

Qy        802 AAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAA 861
              |||  |       |   |||  | :|  ||    :    || | ::    | | |  | |
Db        623 AAAAAATATTTTTTTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAAT 564

Qy        862 AAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATATGATCAACACT 921
              : | :  || |  |  :|||| |   |:  :|  |:  |: | :|:::: |  || : :
Db        563 WTATTWWTTTAAATTATWTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWW 504

Qy        922 GTACTTCGATTTGGGAAACTAGCATATGCAGT 953
```

```
              |: ||  : ||   ||: |    |::  | |
Db        503 WTWTTTTTWATTWATAAWTTTWTTTWWAWATT 472


RESULT 6
CNS0145P/c
LOCUS       CNS0145P                1043 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN11G11 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL103735
VERSION     AL103735.1  GI:5615346
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1043)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES              Location/Qualifiers
     source           1. .1043
                      /organism="Drosophila melanogaster"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:7227"
                      /clone="BACN11G11"
                      /clone_lib="DrosBAC"
                      /plasmid="pBeloBAC11"
                      /note="end : T7"
ORIGIN

  Query Match             5.5%;  Score 55.2;  DB 10;  Length 1043;
  Best Local Similarity   34.6%;  Pred. No. 0.22;
  Matches  166;  Conservative   66;  Mismatches  248;  Indels    0;  Gaps    0;

Qy        431 TTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATCTATTG 490
              | ::  | :|:|||||| || || |: :   : |  :     |  |  |          |
Db       1031 TAWWATTAWTWAATAATTATATAWAWKDTATDGTAGRGGGGGGGGGGGGGGGGGGGGGG 972

Qy        491 GTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGGGATGCCATTG 550
              | ||| | ||| : : : :   || |||  |  |      |   :   |: : :   :
Db        971 GGGGTTTTTTTKKKKKKKKTTTTTTTTTTTTTTTTTTTATATWTTTTDTRTKTTWTDAT 912

Qy        551 ATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGCGTAT 610
              : :|  |:      |:| ||      ||  | |  |  | | ||| |  |  |     : :
Db        911 TWGKTTTTKGTATRTDTKTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTWTW 852

Qy        611 TTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATATGCCGAGGAG 670
              |: ||:          |       : |::| | |  ||       |  ::::::    :  :
Db        851 TWTATWTWWATATATWTARRGGRAADDAAAAAAAAAATTWTTWTWWWWWWWWTTWWWWWWW 792

Qy        671 ATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCAT 730
              ::  | :|  |||  | || |||  |||    | |  ||    || |   |: | |  ||
Db        791 WWTTTTWTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANTATNTWAAATATAT 732

Qy        731 ATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAA 790
              |  | :|| :|  :|   |  | || |  |||||||| | |: :|:|  |  || | :|:
Db        731 AAATTWAAAWATAWATATAAATATAWATAAAAAATATTTATWAWAWAWWATAATATAWAW 672

Qy        791 ATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCATGTTATGGCA 850
              |    :   :|:|| | | |   |  ::|:||| |: : :    |    :   ||:
```

```
Db        671 AAAAAWAWAWAWAAAAWATATATTTWWWAWATAAAWAWAWWWAAAAAAAWATATAWATAT 612

Qy        851 ATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATA 910
              :  | ||  | ||| | || ||| : | ||||||  |:|  ::  |:   |||:||| |
Db        611 TWTTTWATATATAAAAATATATATAKTATTTTTATATAWATAWWTATWTATATAWAAAAA 552


RESULT 7
CNS000D1
LOCUS       CNS000D1               1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR01J16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065414
VERSION     AL065414.1  GI:4938827
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR01J16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.5%;  Score 55;  DB 10;  Length 1101;
  Best Local Similarity   14.7%;  Pred. No. 0.24;
  Matches   70;  Conservative  212;  Mismatches  194;  Indels    0;  Gaps    0;

Qy        335 AAGCAGATATATATTTATGTGACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAA 394
              ||  |||     ::::|::::| |:| |  | :: |  : :::|:: ::::::: :|::|
Db        595 AAAWAGARGATNDDKBAKDWKAAADAKAKAAAWHAGTWWTKKDTKKKKKKKRKWGKAKRA 654

Qy        395 ACCCTCATCGTTCTTGGCTCAGAACAGATGATGAACTTTATGTGTTTAATAATAATTTAG 454
              |   | :|  |: :| : :  :|| ::|:: : :   : |   |: ||  :|: ::::::
Db        655 ATDDTAWTATTWKKTAKKDAGKAAKWKADRTKAWKGAKGAGARGKATAKARAKKRWKDDK 714

Qy        455 ATAGGGAAATCTATCTCAGTAGATGCTTATCTATTGGTGGTGTATTTAGCTATCTAAGTT 514
              :| :::   | | : |   :: ||: :::  :  | |  ::::|   :|  :  ::::|
Db        715 DTKKKKTKTTATKKKTTTTKRAATDKWKWRKKKATAGAWDWDDARRGRGGRWGDKRRRTG 774

Qy        515 CTTTAATAGTAAAAAAAGAACGATGGGATGCCATTGATTTTGATGCGTCCTATATTGGCA 574
              : :: |:     : ::  | :|:: :|     ::: : :| : ::  ::|:  :: |
Db        775 ARGDKDKTGKKKKTGDKRDKGGGDGDRGDGGAGGGDRDRDARARRARWDTKDTRGAKDKA 834

Qy        575 CTTCCTATCCTCATGTATTTATCATGATGAGCGTATTTAATACGCCAGGGTGCCTTTTGC 634
```

```
              ::   ::  :   :|:: :   :: :  :: ::: :  : : :  : :::: :::::
Db        835 DKKKAGRKRKWWTRKTWKRDTADWKAWAAWRWRWRRDRKDRRGRRDRWRRWRGADKKWKW 894

Qy        635 ATTATATATCAAAACCACTCGTAATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAG 694
              ::::||:::    :  :    || :::: ::|:::: ::: : :|::   :  :  ::
Db        895 DDDDTAKDDWTKKDTDWTAGATAGWDDDKAKDGDWKRGTDWRTWKRTWKGWGRKGDKGRK 954

Qy        695 GAAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTT 754
              | : |   :  :  :::  :: :||: ::|::  |:| :::::: ::   : :: : ::
Db        955 TAGRAGDGKDKKKGKDKDRTDDAKATDGKDTWRDTTDTAKRRRRKRRRRAGRGRDRKGDK 1014

Qy        755 ACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGA 810
              :  |:::   : : ::::::: ::  : : :|:||: ::: :  :||| |:|:|
Db       1015 KKRWRARWRKGKKKKTKDKDDDDGKRWGRAKAKADAAKAKDDADAADAAAWAKAKA 1070


RESULT 8
CNS0161D
LOCUS       CNS0161D               1225 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106171
VERSION     AL106171.1  GI:5620504
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1225)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1225
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15C18"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             5.5%;  Score 54.8;  DB 10;  Length 1225;
  Best Local Similarity   30.2%;  Pred. No. 0.27;
  Matches  113;  Conservative   85;  Mismatches  176;  Indels    0;  Gaps    0;

Qy        628 CTTTTGCATTATATATCAAAACCACTCGTAATATGCCGAGGAGATAATGATAGTTTCGAG 687
              :: ::  |: |:| |  :||::   |: :||:|:    |   | |||   |    : |
Db        763 MKMKHMAADAAKAMAMNWAAHHWTTTYTWAAHAYNAAAAATTTAAAATCKCAAAMAMAAM 822

Qy        688 AAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAAT 747
              || ::|  ||   :| : |:  :::: ::: :::: | :: : :: :|||   |  ||
Db        823 AAAMMAAAAACCCMCCVCCKKKKDKKKKKKKKDDKDAAKRARADDRKAAAAAAAAATATA 882

Qy        748 GATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAG 807
              |     | |  |||:::|| : :::| |:  |  : |  ::|:| :||   ::||| |
Db        883 AAAAAAAAAAAAAAAWWWATTYTYYYAYAWAAAYAWATAAWWAWTTWTTAAAWWAAAAAA 942

Qy        808 AGACCATGGTTATATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGA 867
              | |  ::  |:| | | :  |:     |:    | || |   :::  |  : | ||   |
```

```
Db        943 AAAWAWWTTTYAAAAAAWTWTWAAATAAWTTATTTTTTTTTAWWWTTTATWWATAATTTA 1002

Qy        868 GATTTATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATATGATCAACACTGTACTT 927

Db       1003 TWTTTWTTTTTTTTWTAAAAWAAAAAAAAAATATATTTTWTWAWWWTWTWATTTTTATATA 1062

Qy        928 CGATTTGGGAAACTAGCATATGCAGTGAAAAATATTACCGTGCTTAAGAATTTTACTAAA 987

Db       1063 TWTAYTTTTATATTMWAWAWWATWWTTWTWTWTTWTTWWWTWWWTAATATTTTTWTTTTA 1122

Qy        988 CGGATAATTAAGTA 1001

Db       1123 TWTAWAATWATATA 1136


RESULT 9
CNS015W0
LOCUS       CNS015W0                 922 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15B24 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL105978
VERSION     AL105978.1  GI:5619386
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 922)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .922
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15B24"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             5.4%;  Score 54.4;  DB 10;  Length 922;
  Best Local Similarity   33.7%;  Pred. No. 0.32;
  Matches  136;  Conservative   66;  Mismatches  202;  Indels    0;  Gaps    0;

Qy        591 ATTTATCATGATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACC 650

Db         58 AKTTACAGGTAAAAWMTTAMAAAGAMATTRATGCTCCGTTTAAAAAAAAAWAWCTATTCC 117

Qy        651 ACTCGTAATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAAT 710

Db        118 CCAAATWAWATGAAAWAWACAAARATTTAATWAMATWAATRCAAGTGTTCCCTTWCAWAT 177

Qy        711 TTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATC 770

Db        178 CWAAATCCTTCTGTGATAAMAMTTAMAATGCGGTMAAAMCTTCAAAMATAAAARCCGAMC 237

Qy        771 TTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTT 830

Db        238 TTRAWAWMAMAWAWMTWAWMMTGWATWMATWAARGMAMGMWKTTTGTAWTAAWAWAMMAM 297
```

```
Qy        831 GGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAA 890

Db        298 CCCARAARAAMTGAAATCGAAAWMAAAWWMGTTMWAMMAATWWCGTWAAAAMAMCTAMAG 357

Qy        891 GCTAGGTTGTAATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGC 950

Db        358 AMMAGMMTWTGMGAAAMAMMAAAMWCRAAAARACAAMWAAAAAARAAAAAAARATMAAAG 417

Qy        951 AGTGAAAAATATTACCGTGCTTAAGAATTTTACTAAACGGATAA 994

Db        418 GAAAAWWAGGATTAAAGAGAAAAAWRAGAATWMCAMMGAAAAAA 461


RESULT 10
CNS017ZQ/c
LOCUS       CNS017ZQ                 1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37P08 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108704
VERSION     AL108704.1  GI:5629008
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN37P08"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             5.4%;  Score 54.4;  DB 10;  Length 1101;
  Best Local Similarity   33.2%;  Pred. No. 0.33;
  Matches  104;  Conservative   73;  Mismatches  135;  Indels    1;  Gaps    1;

Qy        669 AGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGC 728

Db       1100 ATWTAWWWATWAWWWAWWAWATATATATATATATAWATATAWAAWTAWATAWTATATATA 1041

Qy        729 ATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGA 788

Db       1040 TATTATATAAWATAATWATAWWWWWTWAAWTATATATATATAWWATATATTTWTATTTAT 981

Qy        789 AAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCATGTTATGG 848

Db        980 WTATATWTTATWRATATWTATA-TATAWATATAWATAWWWWWAATTWTTTWAWWWWAWAT 922

Qy        849 CAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAA 908

Db        921 ATATWAWTATWATAATATATAWWATAAAVAARARARAMAKWWATAWATWWTAAAATATAA 862
```

```
Qy        909 TATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATATTACCGT 968
              |:| :|  |:|  :|: :::  ||:    | |  : :|||: |  |:|::|:::|:  |
Db        861 TWTWWTATABATMKTMNHKYMKTTYTTAKAYCAKRKRTATKAAAAGRADMAKWDTWTWTT 802

Qy        969 GCTTAAGAATTTT 981
                |:::| :| :|
Db        801 TTTYRMGTRTGKT 789


RESULT 11
CNS0106X
LOCUS       CNS0106X                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098595
VERSION     AL098595.1  GI:5610206
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN03K20"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             5.4%;  Score 54.2;  DB 10;  Length 1101;
  Best Local Similarity   19.2%;  Pred. No. 0.36;
  Matches   98;  Conservative  203;  Mismatches  209;  Indels    1;  Gaps    1;

Qy        402 TCGTTCTTGGCTCAGAACAGATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGA 461
              |::|| :|   |:   |    || |: | |:    |||:||   :||||:||::||: ::
Db        586 THKTTTKTWTTTMTWKAKGTKTGTTDKAKTKATMATGTKTATRTWTAATKTAKRTADTKD 645

Qy        462 AATCTATCTCAGTAGATGCTTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAAT 521
               :|       | : :| : ::   :|:: ::  : :|:::|       ||:|   ::|::
Db        646 TRTAATATAKATKKKARKDKDTAKDAWKWDKATKAKAKKKAAKAKAKAAAKTATGKDARK 705

Qy        522 AGTAAAAAAAGAACGATGGGATGCCATTGATTTTGAT-GCGTCCTATATTGGCACTTCCT 580
              |:  | | | ::|  :  : ||     :|  ::||: : :   :|:| ::: :  :  :
Db        706 AKAKAKAKATKRAKARAWDTATWTATAADAAADKGAKDKAKAKAKADARDKRRWDAKKDK 765

Qy        581 ATCCTCATGTATTTATCATGATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATA 640
              ::  :    :| ::|: : :: :|   :|: :||:| :   :|:::  :: ::  ::::|
Db        766 RKRAKAKAAKAADDADADAKDAKAADAKAKADAADADDGRGDGKKKRAKDRKKKKKKDKA 825

Qy        641 TATCAAAACCACTCGTAATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGG 700
              : :   |       :  |:| |  :|::::| : ::|:|:|: :  :   |:::::  || :
Db        826 WGDKKKAKKDKAAAKAKADAAGDAKAKRRRAGDKDKAKADAKAAKKAKAKDDDDAKAATK 885

Qy        701 CCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTA 760
                  :: :| ::: ::::: :  |:|::  : ::::: :::::: : ::  :::: 
Db        886 AKKATKAKKDKAKAKKKKKKKDKDAKAKAKAKDAKDKDDKDDDKDDKKDAKADKKKWKKD 945

Qy        761 AAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTAT 820
              :| ::|: : :::: |  :::  :: :| |: : : :     | |:|:|:|  ::   :|:
Db        946 RAKDKAKKDDDKKDKAKDKDKDKKDADAKAWAKAKADADAKAKAKAKAKAADDDAAADAK 1005

Qy        821 ATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAAT 880
              |:| :|  : :   ::   |: : : :  :|:|: :| ::|| |  : :  :|: : :::
Db       1006 AKAADAKAKAKAKADDAAKAKAKADAKAADADAKAKAAKRAAKAKADADAAKAKAKADDK 1065

Qy        881 TTTATGCAAAGCTAGGTTGTAATAAAAATAT 911
              :::|:: :|:: :  : ::  |:| | |:|:
Db       1066 DKDAKKDDADKAKKAKAKDAKAKAKAKAK 1096


RESULT 12
AZ522603/c
LOCUS       AZ522603                562 bp    DNA     linear   GSS 07-MAY-2001
DEFINITION  208PbF09 Pb MBN #21 Plasmodium berghei genomic 3', genomic survey
            sequence.
ACCESSION   AZ522603
VERSION     AZ522603.1  GI:13960786
KEYWORDS    GSS.
SOURCE      Plasmodium berghei
  ORGANISM  Plasmodium berghei
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 562)
  AUTHORS   Carlton,J.M.-R. and Dame,J.B.
  TITLE     The Plasmodium vivax and P. berghei gene sequence tag projects
  JOURNAL   Parasitol. Today 16 (10), 409 (2000)
   PUBMED   11006469
COMMENT     Contact: Dame JB
            Dept. of Pathobiology, College of Veterinary Medicine
            University of Florida
            2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
            Tel: 352 392 4700
            Fax: 352 392 9704
            Email: damej@mail.vetmed.ufl.edu
            Seq primer: M13(-20) forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .562
                     /organism="Plasmodium berghei"
                     /mol_type="genomic DNA"
                     /strain="ANKA clone 15cy1 (clone of the ANKA 8417 clone)"
                     /db_xref="taxon:5821"
                     /dev_stage="asexual blood forms"
                     /lab_host="Mus musculus"
                     /clone_lib="Pb MBN #21"
                     /note="Vector: pBluescript SK(+) vector DNA, phagemid
                     excised from lambda ZAP; Site_1: EcoRV; Site_2: EcoRV;
                     Genomic DNA was prepared from asynchronous blood stage
                     forms of the  cloned ANKA isolate of P. berghei grown in
                     laboratory Swiss white mice.  The DNA was purified from
                     contaminating host DNA by Hoechst Dye  33258-CsCl
                     ultracentrifugation and precipitated. Purified DNA was
                     digested with mung bean nuclease in the presence of 36-38%
                     formamide  at 50 C, as described (Vernick, K.D., Imberski,
                     R.B., and McCutchan, T.F.  1988. Nucleic Acids Research
                     16:6883-6896). The ends of the digestion  fragments were
                     polished using T4 DNA polymerase, and the fragments  size
                     selected in the range 500-2000 bp. These were ligated into
                     the  EcoRV-cleaved and dephosphorylated pBluescript SK(+)
                     vector.  Recombinant plasmids were used to transform E.
                     coli XL10-Gold host cells. "
ORIGIN

  Query Match             5.3%;  Score 53.4;  DB 9;  Length 562;
  Best Local Similarity   50.8%;  Pred. No. 0.52;
  Matches  126;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy        658 ATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATT 717
```

```
Db        557 ATTTATCCCGGAAAGAATAATCATTGAAATATAAATGTGTATATATTATACGTTTTGAAT 498

Qy        718 GATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAA 777

Db        497 TAGTTTATTTATTATTTTTTTTTGTTTTATAAANTTTCCTGTTTATATGTATTTTAAAAA 438

Qy        778 CGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATG 837

Db        437 AAACATATTTGAGATTTAATAAATTCAAGGAAACAAAAGTTGTTCACCATTTTTAATTTA 378

Qy        838 GCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGT 897

Db        377 ATATGTTTAGTCCTATACGATGCAAACACTAAATTATATAAATAGTACACATATCGAAAT 318

Qy        898 TGTAATAA 905

Db        317 TTATAGAA 310
```

RESULT 13
CNS0182P
LOCUS CNS0182P 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL108811
VERSION AL108811.1 GI:5629115
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37D10"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : SP6"
ORIGIN

Query Match 5.3%; Score 53; DB 10; Length 1101;
Best Local Similarity 18.1%; Pred. No. 0.65;
Matches 86; Conservative 190; Mismatches 199; Indels 0; Gaps 0;

```
Qy         29 ATAATAGAAAACAATATCTTGAAGAGTTGTTGAATAGTATAAATAATCAGGAAAAATTTA 88

Db        622 AAAAAAAAAAAAAAAGTTKTTAAAAAAAGGGWGKTGKGKAADGDGGAAAADAWRRGRGRGR 681

Qy         89 ATTTAGATATTGAGATATGTATATCAGATAATGCCTCTACTGATGGTACAGAGGAAATGA 148

Db        682 RARAARAAAAADAKRAAAAAAAAAAAAAAAAAAAADAGGKRKWKKGDGKDGGKATKTAAWAK 741

Qy        149 TTGATGTTTGGAGGAACAATTATAATTTCCCAATAATATATCGGCGTAATAGCGTTAACC 208
```

```
Db        742 GRKDWGTATAWWTWDATWADTWKAATDTDAKRAAAGRRKRDARKTARDGGRRARTRRRAW 801

Qy        209 TTGGGCCAGATAGGAATTTTCTTGCTTCAGTATCCCTTGCGAATGGGGATTATTGTTGGA 268

Db        802 AAGGKRARAGARRARAARRAADDRDWDAWAAAAAAAAAAAAAAWTTWRDRWWDWDDWDTRW 861

Qy        269 TATTTGGCAGTGATGATGCTCTTGCGAAAGACTCGTTAGCGATATTACAAACTTATCTCG 328

Db        862 DDDTTAAWWDDARARARRRRRRRRRRRRARRAARRAADDTDTKDRWADATTDKDTTKTWT 921

Qy        329 ATTCTCAAGCAGATATATATTTATGTGACAGAAAAGAGACCGGGTGTGATTTAGTTGAGA 388

Db        922 TDDDDWDKAKRDRWWAAKADGAWKWRDRARDDWAATAKDDDGWKDKWGGRGKRGGKDKKR 981

Qy        389 TTAGAAACCCTCATCGTTCTTGGCTCAGAACAGATGATGAACTTTATGTGTTTAATAATA 448

Db        982 WDKGKTGKKDDDDWDKTWTRDWWWWWTRTKWDWWWDDGRGRGRWTRKRWGAWWRADAWAR 1041

Qy        449 ATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATCTATTGGTGGTGTATTTAG 503

Db       1042 DDTDGKDTRTADKDRKRTDTTKRDGDDWRKDRKDKKRRDKKGDDKTKKDATWDDD 1096
```

RESULT 14
CNS001FB/c
LOCUS CNS001FB 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR04A23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL060732
VERSION AL060732.1 GI:4939397
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR04A23"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

Query Match 5.3%; Score 52.8; DB 10; Length 1101;
Best Local Similarity 41.6%; Pred. No. 0.72;
Matches 123; Conservative 36; Mismatches 132; Indels 5; Gaps 1;

```
Qy        703 AGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAA 762
```

```
                | |  ::|| ||||: || :::    : ||  || ||:|  |:|| : |  || |  |||
Db       1091 AAAAAWWTTATAATWTATAWWWAAATWAATAAAATATWAAATWATAWATAATAAAAAAAA 1032

Qy        763 AATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATAT 822
              ||| ||  |||||  |  ||:| |||| : |  ||  :|||||
Db       1031 TATAAATWATTAAAAATAAAATWTAAAAAWATWAAAWAAAAAATTWAATAATTAWTATAT 972

Qy        823 ACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTT 882
              ||        |    |  |||    :: |   | ||:| |   || |||   ::|
Db        971 TNAAAAAAAAWAAAAAAAAATTAATAAWWAAAAAAAAAAWATAATAATATATAAAWWTAA 912

Qy        883 TATGCAAAGCTAGGTTGTAATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTA 942
              |:|       |:  ||  || || |||| || :: | |  |     || :   |||  :
Db        911 TWTAA-----TWNATTWAAAAAATAATAWAATAWWAAATTAAWAAAAATAWATWAAAAAW 857

Qy        943 GCATATGCAGTGAAAAATATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAA 998
              |::|  | :  :||| ||:      : || | ::||  | ||   ||| :|||
Db        856 NTAWWTATAAWTTWAAAAATWTNTWAWWWAAAAATWWTTTTTTAATATATATWTAA 801


RESULT 15
CNS0162X
LOCUS       CNS0162X               1201 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15O19 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106227
VERSION     AL106227.1  GI:5620779
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15O19"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             5.2%;  Score 52.6;  DB 10;  Length 1201;
  Best Local Similarity   31.6%;  Pred. No. 0.8;
  Matches  111;  Conservative   86;  Mismatches  149;  Indels    5;  Gaps    1;

Qy        547 ATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGC 606
              :::: |||: ||| :|  | : | :  | ::  |:|  : :||| ||  :  :  || |
Db        697 DWWKTTTTDTATGTDTTTTTKTTAKTAAAWWGWTWTTADTDTGTGTTAGWTTDWTTGTGT 756

Qy        607 GTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATATGCCGA 666
              |||| ||    |  | ::||  ||:     |||:: :  | |  :   |  |: :   |:
Db        757 GTATATATRGTGATATKRTGARTTK-----TATWWKWTGAGATKRTARGAGAKKWTRWGW 811

Qy        667 GGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATT 726
              :    ||: ||::  :  | |  :   |:  :|:  : |     |   |:| : : |::

Db        812 RAGTGTARAGAWRWAKARGKGKKKKKTGKKTWGKDKRWAKWGTGTGAKAWTWRRDATRWG 871

Qy        727 GCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTT 786
              :|||  |:|| :  ::::  | :|:: | :| |:: | |  :||     :::   :|
Db        872 WGWTATTATAWATAWTAWRWKTGTAKTDWTAAKAKAWRAAGAGRKTTTKTKTRDDTTAKT 931

Qy        787 GAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCATGTTAT 846
              |:  |||:|::   : ||: : :  ||:  |||  | |  | ||| |  |  || |||:
Db        932 GWGTGTGTKTKKKAGRTAGRARADAGATKTRTATGGATAGDTGTGGGTGGGWGATTTTAK 991

Qy        847 GGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGT 897
              :  ::| ::  |   |:: ::|::::| || ||||: :: |  | ::|||
Db        992 KTGTRKATKKGGGKTTADRADRAKDKRTGTGTATTTWKWKKAGTGGKRGGT 1042


Search completed: May 12, 2006, 03:30:46
Job time : 3130.9 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2006, 19:15:19 ; Search time 137.043 Seconds
(without alignments)
12996.800 Million cell updates/sec

Title: US-09-423-093B-4_COPY_14059_15060
Perfect score: 1002
Sequence: 1 atgcttatatcattttgtat.....ctaaacggataattaagtag 1002

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	52	5.2	1141	3	US-09-806-708B-22	Sequence 22, Appl
	2	49.6	5.0	3480	2	US-08-920-812-11	Sequence 11, Appl
	3	49.6	5.0	3480	2	US-08-920-827-11	Sequence 11, Appl
	4	49.6	5.0	3480	2	US-08-921-177-11	Sequence 11, Appl
	5	49.6	5.0	3480	2	US-08-362-577C-11	Sequence 11, Appl
	6	49.6	5.0	3480	2	US-08-920-828-11	Sequence 11, Appl
	7	49.6	5.0	580073	3	US-08-545-528D-1	Sequence 1, Appli
c	8	48.4	4.8	95255	3	US-09-949-016-17067	Sequence 17067, A
	9	48	4.8	477	3	US-09-134-000C-812	Sequence 812, App
	10	47.4	4.7	640681	3	US-09-790-988-1	Sequence 1, Appli
c	11	46.8	4.7	3001	3	US-09-539-333D-204	Sequence 204, App
c	12	46.6	4.7	640681	3	US-09-790-988-1	Sequence 1, Appli
c	13	46.4	4.6	1141	3	US-09-806-708B-22	Sequence 22, Appl
c	14	46.2	4.6	601	3	US-09-949-016-107722	Sequence 107722,
	15	46.2	4.6	390890	3	US-09-949-016-14720	Sequence 14720, A
c	16	46	4.6	101894	3	US-09-949-016-12005	Sequence 12005, A
c	17	46	4.6	103894	3	US-09-949-016-14450	Sequence 14450, A
	18	46	4.6	251672	3	US-09-949-016-17296	Sequence 17296, A
	19	46	4.6	251682	3	US-09-949-016-11973	Sequence 11973, A
c	20	45.8	4.6	601	3	US-09-949-016-107721	Sequence 107721,
	21	45.8	4.6	601	3	US-09-949-016-186364	Sequence 186364,
	22	45.6	4.6	601	3	US-09-949-016-186365	Sequence 186365,
c	23	45.2	4.5	3680	3	US-09-647-390-15	Sequence 15, Appl
	24	45	4.5	1080	3	US-09-248-796A-3766	Sequence 3766, Ap
	25	44.4	4.4	50000	3	US-09-662-254B-26	Sequence 26, Appl
	26	44	4.4	4604	3	US-09-830-807-22	Sequence 22, Appl
c	27	43.6	4.4	601	3	US-09-949-016-48813	Sequence 48813, A
c	28	43.6	4.4	601	3	US-09-949-016-48985	Sequence 48985, A
c	29	43.6	4.4	601	3	US-09-949-016-49157	Sequence 49157, A
c	30	43.6	4.4	601	3	US-09-949-016-49333	Sequence 49333, A
	31	43.6	4.4	95255	3	US-09-949-016-17067	Sequence 17067, A
	32	43.2	4.3	832	3	US-09-621-976-2813	Sequence 2813, Ap
	33	43.2	4.3	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	34	43	4.3	1826	3	US-09-286-691-11	Sequence 11, Appl
c	35	43	4.3	1826	3	US-09-687-147-11	Sequence 11, Appl
	36	42.8	4.3	2469	3	US-09-248-796A-5814	Sequence 5814, Ap
c	37	42.4	4.2	20674	3	US-09-641-638-651	Sequence 651, App
c	38	42.4	4.2	20674	3	US-10-170-097-651	Sequence 651, App
c	39	42.4	4.2	251769	3	US-09-949-016-13185	Sequence 13185, A
c	40	42.4	4.2	251769	3	US-09-949-016-13186	Sequence 13186, A
c	41	42.4	4.2	266748	3	US-09-949-016-13187	Sequence 13187, A
c	42	42.4	4.2	266748	3	US-09-949-016-13188	Sequence 13188, A
	43	41.8	4.2	33778	3	US-09-596-002-19	Sequence 19, Appl
c	44	41.8	4.2	101472	3	US-09-949-016-15861	Sequence 15861, A
c	45	41.8	4.2	102738	3	US-09-949-016-12447	Sequence 12447, A

ALIGNMENTS

RESULT 1
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 5.2%; Score 52; DB 3; Length 1141;
Best Local Similarity 12.2%; Pred. No. 0.0039;
Matches 120; Conservative 361; Mismatches 493; Indels 11; Gaps 4;

Qy 23 CAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATAGTATAAATAATCAGGAAA 82
Db 98 YHARRWKDMKTAYBMTMTNKWGKTGWRHRYWRWRAMBDTVDHHYVTAMNNAWTTMCMMDK 157

Qy 83 AATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCTCTACTGATGGTACAGAGG 142
Db 158 DDKRTRWWWKKNNNATGWDDDTKYHMWNNNGCBTVTWMVRYKTDRDWSBKRMNYGMBWWK 217

Qy 143 AAATGATTGATGTTTGGAGGAACAATTATAATTTCCCAATAATATATCGGCGTAATAGCG 202
Db 218 NWSYDVTYYWWVWDDMCKRKVRRWVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMAYRRWT 277

Qy 203 TTAACCTTGGGCCAGATAG--GAATTTTCTTGCTTCAGTATCCCTTGCGAATGGGGATTA 260
Db 278 MNNNNNNAKAMCKRAKYWGWNRABVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDHKWW 337

Qy 261 TTGTTGGATATTTGGCAGTGATGATGCTCTTG-CGAAAGACTCGTTAGCGATATTACAAA 319
Db 338 KWSAAMGVYWNNNNNNNWTYKKARHBARWDWVWHSAWKKWHANAAHYSRKKWTBYKRKTM 397

```
Qy        320 CTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGAAAAGAGACCGGGTGTGATT 379
Db        398 VNNNNGTTMWKRMWAWYWKMDMDWBGTYNNNNNGGRTYYGWTKNKKMWTYYKWKANNCKW 457

Qy        380 TAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACAGATGATGAACTTTATGTGT 439
Db        458 RAWDHKTCTHNNTTWWKMKTYWNNCYWKSMTNGKSHRBAAAVYTWYMWWWRRYAHANNNN 517

Qy        440 TTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATCTATTGGTGGTGTAT 499
Db        518 WDYWWKACTWYKYBVCSKWWNNYAAWYTKSSWNYTSRYYRWKTNNSWRWRSDTRSMGRAN 577

Qy        500 TTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGGGATGCCATTGATTTTGATG 559
Db        578 NYARABHYGYKWNTRWWBWSHTWBHBRAGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGG 637

Qy        560 CGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGCGTATTTAATACGC 619
Db        638 SNNNNNNNNNNNNNNNATCARDDYYAASRW-----YAMANAKWYYYKBAANNAYYTHAN 692

Qy        620 CAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATATGCCGAGGAGATAATGATA 679
Db        693 NWWGCWNNATDTRRTMWKNNNNNNAGTWKNNNNNNAKNASAAKNYAAAAVKAAKKHWRWA 752

Qy        680 GTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAAT 739
Db        753 NKWAMRGWHADAAABTTDKRNNGAYTKYTTTNNNNTYRGVVTNTAARDGWANNNNNNNNN 812

Qy        740 TAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGC 799
Db        813 NNNNNGWSDMWVTWWAYANYGTNNNNNNNNNNAYAWWTNKWYYTTDDRWRBAYTNNNNNN 872

Qy        800 TAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATG 859
Db        873 RMAYYGAYADDYAYYMSDTCDAWMKWDATKMNNATTYNRGTAWRTNNNNNNNMTMKTKYYB 932

Qy        860 AAAAAAGA---GATTTATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATATGATCA 916
Db        933 HAAWNNNNNNGKMCTAHTWWVCKATKTTKGCWMNCTTTCRKYKNNCTWYTWMTTTRTTWY 992

Qy        917 ACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATATTACCGTGCTTAAGA 976
Db        993 AATRWKTNNATGSMTRCNATGWKNNNYWTGWKTRWTAYRMATRWMKAWWKVMATGSWNTN 1052

Qy        977 ATTTTACTAAACGGATAATTAAGTA 1001
Db       1053 SYARWAYKTRAYKGWYYNACAWRWR 1077
```

```
RESULT 2
US-08-920-812-11
; Sequence 11, Application US/08920812
; Patent No. 5763188
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,812
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3480 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Enterococcus faecalis
;      STRAIN:  Clinical Isolate S2-7
US-08-920-812-11

  Query Match             5.0%;  Score 49.6;  DB 2;  Length 3480;
  Best Local Similarity   64.8%;  Pred. No. 0.02;
  Matches   92;  Conservative    0;  Mismatches   44;  Indels    6;  Gaps    1;

Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
Db       2797 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 2856

Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
Db       2857 GTAT------TCTTTTACAAGTCAGTGAAGAAGATGAACTAATTATTTCAGATGATGGTT 2910

Qy        125 CTACTGATGGTACAGAGGAAAT 146
Db       2911 CTACTGATCATACGTTGGAAAT 2932
```

```
RESULT 3
US-08-920-827-11
; Sequence 11, Application US/08920827
; Patent No. 5770375
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,827
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
```

```
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3480 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Enterococcus faecalis
;      STRAIN:  Clinical Isolate S2-7
US-08-920-827-11

  Query Match             5.0%;  Score 49.6;  DB 2;  Length 3480;
  Best Local Similarity   64.8%;  Pred. No. 0.02;
  Matches   92;  Conservative    0;  Mismatches   44;  Indels    6;  Gaps    1;

Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
              | || ||| |||||||| | || |||||| || || ||||||| |  ||    ||  |||
Db       2797 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 2856

Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
              ||||      ||    | || | | |  |||   |||  ||  ||| ||||| ||| |
Db       2857 GTAT------TCTTTTACAAGTCAGTGAAGAAGATGAACTAATTATTTCAGATGATGGTT 2910

Qy        125 CTACTGATGGTACAGAGGAAAT 146
              ||||||||  |||   ||||||
Db       2911 CTACTGATCATACGTTGGAAAT 2932


RESULT 4
US-08-921-177-11
; Sequence 11, Application US/08921177
; Patent No. 5798211
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/921,177
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3480 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Enterococcus faecalis
;      STRAIN:  Clinical Isolate S2-7
US-08-921-177-11

  Query Match             5.0%;  Score 49.6;  DB 2;  Length 3480;
  Best Local Similarity   64.8%;  Pred. No. 0.02;
  Matches   92;  Conservative    0;  Mismatches   44;  Indels    6;  Gaps    1;

Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
              | || ||| |||||||| | || |||||| || || ||||||| |  ||    ||  |||
Db       2797 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 2856

Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
              ||||      ||    | || | | |  |||   |||  ||  ||| ||||| ||| |
Db       2857 GTAT------TCTTTTACAAGTCAGTGAAGAAGATGAACTAATTATTTCAGATGATGGTT 2910

Qy        125 CTACTGATGGTACAGAGGAAAT 146
              ||||||||  |||   ||||||
Db       2911 CTACTGATCATACGTTGGAAAT 2932


RESULT 5
US-08-362-577C-11
; Sequence 11, Application US/08362577C
; Patent No. 5807673
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,577C
;      FILING DATE:  27-MAR-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3480 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Enterococcus faecalis
;      STRAIN:  Clinical Isolate S2-7
US-08-362-577C-11

  Query Match             5.0%;  Score 49.6;  DB 2;  Length 3480;
  Best Local Similarity   64.8%;  Pred. No. 0.02;
  Matches   92;  Conservative    0;  Mismatches   44;  Indels    6;  Gaps    1;

Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
              | || ||| |||||||| | ||| |||||| || || ||||||| |  ||     ||  |||
Db       2797 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 2856

Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
              ||||      ||    | || | | |  |||   |||  ||  ||| |||||| |||  |
Db       2857 GTAT------TCTTTTACAAGTCAGTGAAGAAGATGAACTAATTATTTCAGATGATGGTT 2910

Qy        125 CTACTGATGGTACAGAGGAAAT 146
              ||||||||  |||   ||||||
Db       2911 CTACTGATCATACGTTGGAAAT 2932


RESULT 6
US-08-920-828-11
; Sequence 11, Application US/08920828
; Patent No. 5853998
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,828
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3480 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Enterococcus faecalis
;      STRAIN:  Clinical Isolate S2-7
US-08-920-828-11

  Query Match             5.0%;  Score 49.6;  DB 2;  Length 3480;
  Best Local Similarity   64.8%;  Pred. No. 0.02;
  Matches   92;  Conservative    0;  Mismatches   44;  Indels    6;  Gaps    1;

Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
              | || ||| |||||||| | || |||||| || || ||||||| |  ||     ||  |||
Db       2797 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 2856

Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
              ||||      ||    | || | | |  |||   |||  ||  ||| |||||| |||  |
Db       2857 GTAT------TCTTTTACAAGTCAGTGAAGAAGATGAACTAATTATTTCAGATGATGGTT 2910

Qy        125 CTACTGATGGTACAGAGGAAAT 146
              ||||||||  |||   ||||||
Db       2911 CTACTGATCATACGTTGGAAAT 2932


RESULT 7
US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
;  APPLICANT: Fraser et al.
;  TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragments
; Patent No. 6537773
;  TITLE OF INVENTION:  Thereof, and Uses Thereof
;  FILE REFERENCE: PB193P1
;  CURRENT APPLICATION NUMBER: US/08/545,528D
;  CURRENT FILING DATE:  1995-10-19
;  PRIOR APPLICATION NUMBER: US 08/488,018
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: US 08/473,545
;  PRIOR FILING DATE: 1995-06-07
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 580073
;   TYPE: DNA
;   ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

  Query Match             5.0%;  Score 49.6;  DB 3;  Length 580073;
  Best Local Similarity   46.0%;  Pred. No. 0.077;
  Matches  209;  Conservative    0;  Mismatches  239;  Indels    6;  Gaps    1;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||    ||||  |   ||| || ||||| ||| |   |  || | |||| |     |
Db      67651 ATGAAATTATCTGTAATTATACCTACTTACAATTGTGCATCATTTATTGAAAAAGCAATT 67710

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              |||   ||   |||  |   |    | |||| || || ||  | |  ||    ||| ||
Db      67711 AATTCAATTGTTAAAAATAGACCTAATGATTTGGAAATAGAAGTTTTAATTATTGATGAT 67770

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              |  || | |||   |||  |  || | ||| |          || |   ||||     |
Db      67771 GGATCAATTGACAATACTAACAAAGTTATTAAGAAAATTCAAGACCAAATTAATAATTTA 67830

Qy        181 ATAATATATCGGCGT------AATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCT 234
              |   |  |      |      | ||  | ||||   ||    | ||  |||| | ||
Db      67831 ACTTTGCAGTATTTTTACAAAAGTAATGGTAACTGGGGTAGTGTTATTAATTATGTTAGA 67890

Qy        235 TCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCG 294
              || || || ||||| |          |||| |  ||||||||   | || |
Db      67891 AACAATAAACTAGCAAAAGGGGAATGAGTAACAGTATTGGATAGTGATGACATTTTTTCA 67950

Qy        295 AAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGT 354
              ||| |  |  |  | || || ||||  ||| | |    ||| | ||       |   |
Db      67951 AAAAAAACAATTTCTATTTTTCAAAAATATGCCCAAAAACAAAGATATGATGCGATTATT 68010

Qy        355 GACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTC 414
                | |  | || |     ||  |  | ||  |||     | |  || |
Db      68011 TTTGACTACTATAAATGCTGAAAAAAGTTTTTGTGAAAAATTCCTACCTATGCAAGGTTT 68070
```

```
Qy        415 AGAACAGATGATGAACTTTATGTGTTTAATAATA 448
              |||| |||   | ||  | |  ||   ||  | |
Db      68071 AGAAAAGAAATTAAAGGTGAATTGAAAAAACAAA 68104


RESULT 8
US-09-949-016-17067/c
; Sequence 17067, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17067
;   LENGTH: 95255
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(95255)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17067

  Query Match             4.8%;  Score 48.4;  DB 3;  Length 95255;
  Best Local Similarity   42.9%;  Pred. No. 0.093;
  Matches  241;  Conservative    0;  Mismatches  321;  Indels    0;  Gaps    0;

Qy        372 GTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACAGATGATGAACT 431
              ||   |||   ||  |  || ||    | || ||| |    |    | |    || |
Db      43125 GTTATATTATATTTTGTATATAATATATAATTGTTATATTTTGTATATAATATATAATTA 43066

Qy        432 TTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATCTATTGG 491
              | | || ||||||||   |||   ||| |||   |    ||  ||||  ||
Db      43065 CATATTATATATTAATAATTATTATATATTAATATATAATATATAATAATTATTAATATA 43006

Qy        492 TGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGGGATGCCATTGA 551
              |  | |||| |  |||   |  | | || || |||  |||  |  ||   || | ||
Db      43005 TAATATATTAATATATAATAAATATATATTAATAATTAAAATATAATATAATTCTATATT 42946

Qy        552 TTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGCGTATT 611
              | |  ||  |   ||||    | |   |    ||   | |||  |              | ||
Db      42945 ATATTTTGTATATAATATAAAATCATATATTATATATAATTCTATATTATGTTATATTTT 42886

Qy        612 TAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATATGCCGAGGAGA 671
              |||    |      | | |  | ||||||     |       |  | |    |  | |
Db      42885 GTATATAATATATAATCATATATAATATATAATTCTATATTATATTTTGTATATAATATA 42826

Qy        672 TAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATA 731
              |||| |||  ||  | |  |    | |  |       ||  ||  | ||| | ||  ||
Db      42825 TAATTATATATTATATATAATTCTATATACTTTGTATATAATATATAATTCTGTTATATT 42766

Qy        732 TTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAA 791
              ||  | || |  ||    | | |    | | | | | | ||| |      | ||  |
Db      42765 TTGTATATAATATACAATTCTGTTATATTACATTTTGTATATAATATACAATTCTGTTAT 42706

Qy        792 TGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCATGTTATGGCAA 851
              |||   || || | | |       |  ||||      |          |  |||
Db      42705 ATTTTGTATATAATATACAATTCAATGTTATATTTTGTATATACCATTGTTATATTTTGT 42646

Qy        852 TAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATAT 911
              | ||  | ||       |  | || || | ||  | |  |  ||| || || |  |
Db      42645 ATATAATATATAATTGTTATATTTTGTATATAATATATAATTCTGTTATATTATATTTTG 42586


Qy        912 GATCAACACTGTACTTCGATTT 933
              ||  |   | || |||   ||
Db      42585 TATATAATATATAATTCCTGTT 42564


RESULT 9
US-09-134-000C-812
; Sequence 812, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;  TITLE OF INVENTION:  ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-032
;  CURRENT APPLICATION NUMBER: US/09/134,000C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/055,778
;  PRIOR FILING DATE: 1997-08-15
;  NUMBER OF SEQ ID NOS: 6812
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 812
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
US-09-134-000C-812

  Query Match             4.8%;  Score 48;  DB 3;  Length 477;
  Best Local Similarity   64.1%;  Pred. No. 0.028;
  Matches   91;  Conservative    0;  Mismatches   45;  Indels    6;  Gaps    1;

Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
              | || ||| |||||||| | || |||||| || || ||||||| |  ||    ||  |||
Db          5 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 64

Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
              ||||       ||    | || | | |  |||   |||  ||  ||| |||||| |||  |
Db         65 GTAT------TCTTTTACAAGTCAGTGAAGAAGATGAACTAATTATTTCAGATGATGGTT 118

Qy        125 CTACTGATGGTACAGAGGAAAT 146
              |||||||  |||   ||||||
Db        119 ATACTGATCATACGTTGGAAAT 140


RESULT 10
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             4.7%;  Score 47.4;  DB 3;  Length 640681;
  Best Local Similarity   51.7%;  Pred. No. 0.27;
  Matches  108;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        703 AGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAA 762
              | |  ||||  ||||   | ||||  |  | | |   || | |  | ||     | ||
```

```
Db     325630 AAAATAATTACAATTCTATCTATTATAGGTATTAGTATAACGACAGCTTCACTAATTATT 325689

Qy        763 AATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATAT 822

Db     325690 ATTACATCTATAATAAATGGATCTGAGAAAAATTTTGAAAAAAATATTTTATCATTTATT 325749

Qy        823 ACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTT 882

Db     325750 CCACATTTAATAATAACTAATAAAAATCAATGTATTAAAAAAGATCAATTTCCTGAAAAT 325809

Qy        883 TATGCAAAGCTAGGTTGTAATAAAAATAT 911

Db     325810 ATTTTAAAATTAAATAATATTAAAAATAT 325838


RESULT 11
US-09-539-333D-204/c
; Sequence 204, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE: GENSET.047AUS
;  CURRENT APPLICATION NUMBER: US/09/539,333D
;  CURRENT FILING DATE:  2000-03-30
;  PRIOR APPLICATION NUMBER: US 60/126,903
;  PRIOR FILING DATE: 1999-03-30
;  PRIOR APPLICATION NUMBER: US 60/131,971
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/132,065
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/143,928
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 60/145,915
;  PRIOR FILING DATE: 1999-07-27
;  PRIOR APPLICATION NUMBER: US 60/146,453
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/146,452
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/162,288
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: US 09/416,384
;  PRIOR FILING DATE: 1999-10-12
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: Patent.pm
; SEQ ID NO 204
;   LENGTH: 3001
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 1501
;   OTHER INFORMATION: 99-26267-524 : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 1481..1500
;   OTHER INFORMATION: 99-26267-524.mis1,
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 1502..1521
;   OTHER INFORMATION: 99-26267-524.mis2,  complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 983..1002
;   OTHER INFORMATION: upstream amplification primer
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1553..1573
;   OTHER INFORMATION: downstream amplification primer, complement
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 1489..1513
;   OTHER INFORMATION: 99-26267-524  probe
US-09-539-333D-204

  Query Match             4.7%;  Score 46.8;  DB 3;  Length 3001;
  Best Local Similarity   49.2%;  Pred. No. 0.09;
  Matches  123;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy        710 TTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATAT 769

Db       2198 TATATATGAATATAATGTATAATGTAAAATTATATACATTATTAAACTATAAAATTATAT 2139

Qy        770 CTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTT 829

Db       2138 AGTAAAATTATATAATGTATATTTAATATTTTATTTATATATAATGTATAAATATACATA 2079

Qy        830 TGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAA 889

Db       2078 TTTATATGTAATGTATATATATTATTTATACACATATATACACATATATAATATATACAT 2019

Qy        890 AGCTAGGTTGTAATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATG 949

Db       2018 ATATTCCATTTTATATATATACACACAAATATGATCTGGGGAAAAGGTTTCGTGCATACA 1959

Qy        950 CAGTGAAAAA 959

Db       1958 CACACACAAA 1949


RESULT 12
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             4.7%;  Score 46.6;  DB 3;  Length 640681;
  Best Local Similarity   51.4%;  Pred. No. 0.42;
  Matches  133;  Conservative    0;  Mismatches  124;  Indels    2;  Gaps    1;

Qy        745 AATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAA 804

Db     215416 AAATGTTTTAACATTTTTTGCCACCATTTAAAAAGAACCCTATTATGAATATTGATAAAA 215357

Qy        805 GAGAGACCATGGTTATATACAACTTTGGCTATGGCATGTTATGGCAATAG--TGATGAAA 862

Db     215356 TAATGAATTTATTCTCAAAAAGAACTGTTTATGACTTGATAAATAAATAGAATTATATAA 215297

Qy        863 AAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATATGATCAACACTG 922

Db     215296 AAAAATATTTAATTATATTTTATGTTAACATAGTTTATTATAATAATTTCTTTACAAATA 215237

Qy        923 TACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATATTACCGTGCTTAAGAATTTTA 982
```

```
Db     215236 AAATAAGAAAATTATATTTTTAATATACTTTAAAAAAAATATTTTATTTAAAATTTTTCT 215177

Qy        983 CTAAACGGATAATTAAGTA 1001
              || |     ||| ||||
Db     215176 TTATATTTGATATTTAGTA 215158


RESULT 13
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             4.6%;  Score 46.4;  DB 3;  Length 1141;
  Best Local Similarity   12.5%;  Pred. No. 0.087;
  Matches  123;  Conservative  333;  Mismatches  521;  Indels    9;  Gaps    2;

Qy         10 TCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATAGTATA 69
              :::::|   : :|:: :  :|: |:: | :: || :::: :    |:  | :  :  |::
Db       1078 WYWYWTGTNRRWCMRTYAMRTWYTRSNANWSCATKBMWWTMKWYATKYRTAWYAMWCAWR 1019

Qy         70 AATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCTCTACT 129
                ::    | |    |  | ::|   :   | |:: ::|   ::   || |  :  :
Db       1018 NNNMWCATNGYAKSCATNNAMWYATTRWAAYAAAKWARWAGNNMRMYGAAAGNKWGCMAA 959

Qy        130 GATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCAATAATATAT 189
               || |:: |:  | :       :| :  :  | :|         :|  :  |||  ::  :
Db        958 MATMGBWWADTAGKMCNNNNNNWTTDVRRMAMKAKNNNNNNAYWTACYNRAATNNKMATH 899

Qy        190 CGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTATCCCTTGCG 249
               :  : |: : :| ::  | ::  :      :  :::::    ::  ::|
Db        898 WMKWTHGAHSKRRTRHHTRTCRRTKYNNNNNNNARTVYWYHHAARRWMNAWWTRTNNNNNN 839

Qy        250 AATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGACTCGTTAGCG 309
                  : |  :: :::  :                 | |::||       :      |
Db        838 NNNNACRNTRTWWABWKHSWCNNNNNNNNNNNNNNNTWCHYTTANABBCYRANNNNAAARM 779

Qy        310 ATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGAAAAGAGACC 369
               |     :::|  | |:|   :: ::   ::::  |:: |||:  |            :
Db        778 ARTCNNYMHAAVTTTHTDWCYKTWMNTWYWDMMTTMBTTTTRNMTTSTNMTNNNNNNMWA 719

Qy        370 GGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACAGATGATGAA 429
                    ::| :  |   :  ::      :|  ||:  ::        :  | : :::
Db        718 CTNNNNNNMWKAYYAHATNNWGCWWNNTDARRTNNTTVMRRRWMTNTKTRWYSTTR-RHH 660

Qy        430 CTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATCTATT 489
               :|                        |:   :::|:: :   ::: : :: || |: : :
Db        659 YTGATNNNNNNNNNNNNNNNNSCCTCTRMMTMRWTMKGDGMTVRKKVKWRDTTCTYVDVW 600

Qy        490 GGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGGGATGCCATT 549
               :  :: :| : :  ::|:|:  |:|: :|:   ::   |:: ::::|   :      |
Db        599 ADSWVWWYANWMRCRDVTYTRNNTYCKSYAHSYWYWSNNAMWYRRYSARNWSSMARWTTR 540

Qy        550 GATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGCGTA 609
                 ::  |  :::: |::: :     |:  : :  : : :: ||  :::  |:::  |
Db        539 NNWWMSGBVRMRWAGTMWWRHWNNNNTDTRYYWWWKRWARBTTTVYDSMCNAKSMWRGNN 480

Qy        610 TTTAATACGCCAGGGTGCCTTTT--------GCATTATATATCAAAACCACTCGTAATAT 661
               :  :  :     : |   ::|::       : |:::   |:|::| ||       |  ::
Db        479 WRAMKMWWAANNDAGAMDHWTYWMGNNTMWMRRAWKMMNMAWCRRAYCCNNNNNRACVWH 420

Qy        662 GCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATT 721
               ::  ::    :::: |      :  : :::|::   ::    : ::   ::::::::|:::|
Db        419 KHKMWRWTWKYMWKAACNNNNBKAMYMRVAWMMYSRDTTNTDWMMWTSDWBWHWYTVDYT 360

Qy        722 TTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAG 781
                 : :        ::   ||  ::::: : |: |    |:  | :: | : : | |  ::
Db        359 MMRAWNNNNNNNWRBCKTTSWMWWMDHMNTHCTYGNNTWGSAYBMAAMSMWAAGASNBVT 300

Qy        782 CATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCAT 841
               : : :::  : :| | |       |   :   :|:     : :  :: ::   : :  :::
Db        299 YNWCWRMTYMGKTMTNNNNNNKAWYYRTKTVAWCNNRYYYDTAVWTBKRNYKYCYAYBWY 240

Qy        842 GTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGTA 901
               :   |:::::  :|: : : :: ::   :  :  : :    :: | |:     ::: :
Db        239 YBMYMGKHHWBWWRRABHRSWNMWWVKCRNKYMVSWHYHAMRYBKWABAVGCNNNWKDRM 180

Qy        902 ATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAATA 961
               |:::: ||   :  :::|:  ::::: :| :|  : |  :: :|::: :: :: : :  :
Db        179 AHHHWCATNNNMMWWWYAYMHHMHKKGKAAWTNNKTABRDDHBAHVKTYWYWRYDYWCAM 120

Qy        962 TTACCGTGCTTAAGAATTTTACTAAA 987
               ::  : :: ||::::::::|: :: |
Db        119 CWMNAKAKVRTAMKHMWYYTDRYVSA 94


RESULT 14
US-09-949-016-107722/c
; Sequence 107722, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 107722
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-107722

  Query Match             4.6%;  Score 46.2;  DB 3;  Length 601;
  Best Local Similarity   56.1%;  Pred. No. 0.082;
  Matches   87;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        683 TCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAG 742
              | ||||| |||||||||    | | |||| ||  ||  | |||||||| |||||  | |
Db        390 TGGAGAAAAAAGGAAAGCGTATATGAATATTCTTTACTATTATTGCATCTTTAATGTAAA 331

Qy        743 CTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAA 802
               | |  |||||    |  | |||| |||    || |    | ||  |||| |    |   |
Db        330 TTGAACATTTTATAAAAATAAATCTATGARCAACATATGAATGAGAAAAAGCAAAGACCA 271

Qy        803 AAGAGAGACCATGGTTATATACAACTTTGGCTATG 837
               |  |  |    || | ||  |  |    ||  |||
Db        270 AGAACCGTAGTTGTTAATCAAAGAGCAGGGACATG 236
```

```
RESULT 15
US-09-949-016-14720
; Sequence 14720, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14720
;   LENGTH: 390890
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(390890)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14720

  Query Match             4.6%;  Score 46.2;  DB 3;  Length 390890;
  Best Local Similarity   56.1%;  Pred. No. 0.46;
  Matches   87;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy        683 TCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAG 742
              | ||||| |||||||||    | | |||| ||   ||   | ||||||||| |||||  | |
Db      76795 TGGAGAAAAAAGGAAAGCGTATATGAATATTCTTTACTATTATTGCATCTTTAATGTAAA 76854

Qy        743 CTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAA 802
              | |  |||||    |  | |||| |||     || |   | ||  ||||| |    |   |
Db      76855 TTGAACATTTTATAAAAATAAATCTATGAGCAACATATGAATGAGAAAAAGCAAAGACCA 76914

Qy        803 AAGAGAGACCATGGTTATATACAACTTTGGCTATG 837
              |  |  |    || | ||  |  |    ||  |||
Db      76915 AGAACCGTAGTTGTTAATCAAAGAGCAGGGACATG 76949


Search completed: May 12, 2006, 03:52:03
Job time : 143.043 secs
```

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:31:18 ; Search time 641.566 Seconds
(without alignments)
12915.142 Million cell updates/sec

Title: US-09-423-093B-4_COPY_14059_15060
Perfect score: 1002
Sequence: 1 atgcttatatcatttttgtat.....ctaaacggataattaagtag 1002

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1002	100.0	22080	3	US-09-423-093A-4	Sequence 4, Appli
c	2	51.6	5.1	592	4	US-09-925-065A-651832	Sequence 651832,
	3	50.2	5.0	73334	6	US-10-311-455-2097	Sequence 2097, Ap
	4	50.2	5.0	73334	7	US-10-240-589C-127	Sequence 127, App
	5	49.6	5.0	894	7	US-10-282-122A-27240	Sequence 27240, A
	6	49.6	5.0	2601	3	US-09-960-870-17	Sequence 17, Appl
	7	49.6	5.0	2601	3	US-09-960-858-17	Sequence 17, Appl
	8	49.6	5.0	2601	7	US-10-251-668-17	Sequence 17, Appl
	9	49.6	5.0	580073	6	US-10-205-220-1	Sequence 1, Appli
c	10	49.2	4.9	2012	3	US-09-070-927A-443	Sequence 443, App
	11	47.4	4.7	640681	3	US-09-790-988-1	Sequence 1, Appli
	12	47.2	4.7	3673778	6	US-10-312-841-2	Sequence 2, Appli
	13	47	4.7	10048	7	US-10-433-793-46	Sequence 46, Appl
c	14	47	4.7	3673778	6	US-10-312-841-1	Sequence 1, Appli
c	15	46.8	4.7	3001	6	US-10-147-603-204	Sequence 204, App
	16	46.6	4.7	2431	6	US-10-311-455-2241	Sequence 2241, Ap
	17	46.6	4.7	18997	5	US-10-172-086-18	Sequence 18, Appl
	18	46.6	4.7	18997	6	US-10-311-455-544	Sequence 544, App
	19	46.6	4.7	18997	7	US-10-311-507-32	Sequence 32, Appl
	20	46.6	4.7	18997	8	US-10-480-846-18	Sequence 18, Appl
	21	46.6	4.7	19787	6	US-10-311-455-1424	Sequence 1424, Ap
c	22	46.6	4.7	640681	3	US-09-790-988-1	Sequence 1, Appli
	23	46.4	4.6	9997	6	US-10-311-455-342	Sequence 342, App
	24	46.4	4.6	9997	6	US-10-240-485-38	Sequence 38, Appl
	25	46.4	4.6	12441	3	US-09-423-093A-3	Sequence 3, Appli
	26	46.2	4.6	518360	7	US-10-367-094-125	Sequence 125, App
	27	46	4.6	543	7	US-10-021-323-12208	Sequence 12208, A
c	28	46	4.6	65359	3	US-09-804-472-3	Sequence 3, Appli
c	29	46	4.6	65359	10	US-11-057-169-3	Sequence 3, Appli
	30	46	4.6	337344	8	US-10-388-838-58	Sequence 58, Appl
	31	45.8	4.6	415	4	US-09-925-065A-154135	Sequence 154135,
	32	45.8	4.6	3673778	6	US-10-312-841-1	Sequence 1, Appli
c	33	45.6	4.6	8056	8	US-10-473-126-386	Sequence 386, App
c	34	45.4	4.5	604	4	US-09-925-065A-226132	Sequence 226132,
	35	45.4	4.5	34688	7	US-10-433-793-90	Sequence 90, Appl
	36	45.4	4.5	73334	6	US-10-311-455-2098	Sequence 2098, Ap
	37	45.4	4.5	73334	7	US-10-240-589C-128	Sequence 128, App
	38	45	4.5	1152	6	US-10-032-585-6280	Sequence 6280, Ap
	39	45	4.5	5768	6	US-10-311-455-2136	Sequence 2136, Ap
c	40	45	4.5	19659	6	US-10-311-455-739	Sequence 739, App
	41	45	4.5	29993	8	US-10-482-823-2	Sequence 2, Appli
	42	45	4.5	29993	8	US-10-482-823-4	Sequence 4, Appli
	43	45	4.5	38342	7	US-10-221-714A-471	Sequence 471, App
	44	44.8	4.5	6131	6	US-10-311-455-864	Sequence 864, App
	45	44.8	4.5	19634	8	US-10-473-126-302	Sequence 302, App

ALIGNMENTS

RESULT 1
US-09-423-093A-4
; Sequence 4, Application US/09423093A
; Publication No. US20030018349A1
; GENERAL INFORMATION:
; APPLICANT: Reeves, Peter Richard et al
; TITLE OF INVENTION: Nucleic Acid Molecules Specific for Bacterial Antigens And Uses
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 749670/PAS
; CURRENT APPLICATION NUMBER: US/09/423,093A
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PO6545
; PRIOR FILING DATE: 1997-05-01
; PRIOR APPLICATION NUMBER: PO8162
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 22080
; TYPE: DNA
; ORGANISM: Salmonella enterica
US-09-423-093A-4

Query Match 100.0%; Score 1002; DB 3; Length 22080;
Best Local Similarity 100.0%; Pred. No. 2.5e-201;
Matches 1002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14059 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 14118

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14119 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 14178

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14179 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 14238

Qy        181 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14239 ATAATATATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTA 14298

Qy        241 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14299 TCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGAC 14358
```

```
Qy        301 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14359 TCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGA 14418

Qy        361 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14419 AAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACA 14478

Qy        421 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14479 GATGATGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGC 14538

Qy        481 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14539 TTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGG 14598

Qy        541 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14599 GATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATG 14658

Qy        601 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14659 ATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATA 14718

Qy        661 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14719 TGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGAT 14778

Qy        721 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14779 TTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGA 14838

Qy        781 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14839 GCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCA 14898

Qy        841 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14899 TGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTGT 14958

Qy        901 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14959 AATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATGCAGTGAAAAAT 15018

Qy        961 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002
              ||||||||||||||||||||||||||||||||||||||||||
Db      15019 ATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 15060
```

```
RESULT 2
US-09-925-065A-651832/c
; Sequence 651832, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 651832
;   LENGTH: 592
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-651832

  Query Match             5.1%;  Score 51.6;  DB 4;  Length 592;
  Best Local Similarity   53.5%;  Pred. No. 0.71;
  Matches  108;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        708 AATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATAT 767
              || ||| ||| |  |||| || | ||  ||| |  || |||| |||||  |
Db        496 AAGTTTTATTTACATTATATTTTAGTCTAAGCGAGCAATAGAATTTTGTCTAAAAAAAAC 437

Qy        768 ATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAAC 827
              |   | | ||  ||||| |||||  |||  | | | | | | |||||  ||   || |
Db        436 AATGTATACACCTTCATTTAAAAATACTTTCTTGATAAATAAAAAATGGTAATGATCATC 377

Qy        828 TTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGC 887
              |  | || | ||  || ||      |  | | | | || |||||  ||   | ||
Db        376 TGAGCCTTTAGCTAGTCATAATCTTTTTTTTTATTAGAGGGATTTGCCTTTCTGTTGATG 317

Qy        888 AAAGCTAGGTTGTAATAAAAAT 909
              |  |||  |||  | ||  ||
Db        316 ACTGCTGATTTGCCAGAATGAT 295


RESULT 3
US-10-311-455-2097
; Sequence 2097, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2097
;   LENGTH: 73334
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2097

  Query Match             5.0%;  Score 50.2;  DB 6;  Length 73334;
  Best Local Similarity   45.2%;  Pred. No. 9.4;
  Matches  184;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy        474 TAGATGCTTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGA 533
              |||    |  | ||| |  ||||| ||||| ||  | | |   | |  | ||||  |
Db      36186 TAGTGTTTGTTGTATAGTAGGTGTTTTTAGGTAAATGATTGAATGATAAATAAATTATAT 36245

Qy        534 ACGATGGGATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATT 593
              | |  |||   ||| |      |       | | ||     ||  |         |||
Db      36246 CGGTTATGATAGTATTTAAAAGAAAATAGTTTTTTTTTTGTTTTTTTTGTAAGGAGTAAG 36305

Qy        594 TATCATGATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACT 653
              |   |  |     | | ||  ||     ||       ||  | |||   |    | |    |
Db      36306 GAAATTTTTTTTGGAAGTTTTTATATTAGATATTTTTACGTATTGGTTTAGGATATAGTT 36365

Qy        654 CGTAATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTT 713
                | ||| |   | |  ||  |  |||||        ||||| |      |   |||| |
```

```
Db      36366 ATTGATAAGGGAATGGAATTTTTTTAGTTGGTTTGGAAAAGGTAGTTTGTGTTTAATTAT 36425

Qy        714 AATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTT 773

Db      36426 AAAGATTATTGTATTGGTTATAAAAATAATATATGATTATTATAGGAAATTTATAAATTA 36485

Qy        774 AAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGC 833

Db      36486 ATGAAAAGTATAATAAAATAATTTGGTATTATATAAAGAAAATTGTTGTTAATATTGTGG 36545

Qy        834 TATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAAT 880

Db      36546 TGTTTGTTTTTTAGTTTTGGTTGTATATGTATATGTTTGTGTGTAT 36592


RESULT 4
US-10-240-589C-127
; Sequence 127, Application US/10240589C
; Publication No. US20040076956A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  DNA repair
;  FILE REFERENCE: 5013.1008
;  CURRENT APPLICATION NUMBER: US/10/240,589C
;  CURRENT FILING DATE:  2003-09-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03972
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 148
; SEQ ID NO 127
;   LENGTH: 73334
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-589C-127

  Query Match             5.0%;  Score 50.2;  DB 7;  Length 73334;
  Best Local Similarity   45.2%;  Pred. No. 9.4;
  Matches  184;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy        474 TAGATGCTTATCTATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGA 533

Db      36186 TAGTGTTTGTTGTATAGTAGGTGTTTTTAGGTAAATGATTGAATGATAAATAAATTATAT 36245

Qy        534 ACGATGGGATGCCATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATT 593

Db      36246 CGGTTATGATAGTATTTAAAAGAAAATAGTTTTTTTTTTGTTTTTTTTGTAAGGAGTAAG 36305

Qy        594 TATCATGATGAGCGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACT 653

Db      36306 GAAATTTTTTTTGGAAGTTTTTATATTAGATATTTTTACGTATTGGTTTAGGATATAGTT 36365

Qy        654 CGTAATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTT 713

Db      36366 ATTGATAAGGGAATGGAATTTTTTTAGTTGGTTTGGAAAAGGTAGTTTGTGTTTAATTAT 36425

Qy        714 AATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTT 773

Db      36426 AAAGATTATTGTATTGGTTATAAAAATAATATATGATTATTATAGGAAATTTATAAATTA 36485

Qy        774 AAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTTTGGC 833

Db      36486 ATGAAAAGTATAATAAAATAATTTGGTATTATATAAAGAAAATTGTTGTTAATATTGTGG 36545

Qy        834 TATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAAT 880

Db      36546 TGTTTGTTTTTTAGTTTTGGTTGTATATGTATATGTTTGTGTGTAT 36592


RESULT 5
US-10-282-122A-27240
; Sequence 27240, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27240
;   LENGTH: 894
;   TYPE: DNA
;   ORGANISM: Mycoplasma genitalium
US-10-282-122A-27240

  Query Match             5.0%;  Score 49.6;  DB 7;  Length 894;
  Best Local Similarity   46.0%;  Pred. No. 2.2;
  Matches  209;  Conservative    0;  Mismatches  239;  Indels    6;  Gaps    1;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60

Db          1 ATGAAATTATCTGTAATTATACCTACTTACAATTGTGCATCATTTATTGAAAAAGCAATT 60

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120

Db         61 AATTCAATTGTTAAAAATAGACCTAATGATTTGGAAATAGAAGTTTTAATTATTGATGAT 120

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180

Db        121 GGATCAATTGACAATACTAACAAAGTTATTAAGAAAATTCAAGACCAAATTAATAATTTA 180
```

```
Qy        181 ATAATATATCGGCGT------AATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCT 234
              |   |  |      |      | ||  | ||||   ||     | ||  |||| | ||
Db        181 ACTTTGCAGTATTTTTACAAAAGTAATGGTAACTGGGGTAGTGTTATTAATTATGTTAGA 240

Qy        235 TCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCG 294
                || || || ||||| |          |||| |  |||||||    | || |
Db        241 AACAATAAACTAGCAAAAGGGGAATGAGTAACAGTATTGGATAGTGATGACATTTTTTCA 300

Qy        295 AAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGT 354
              ||| |  |  |  | || || ||||  |||  | |    |||  | ||       |   |
Db        301 AAAAAAACAATTTCTATTTTTCAAAAATATGCCCAAAAACAAAGATATGATGCGATTATT 360

Qy        355 GACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTC 414
                | | |   | ||  |      ||  |   | ||  |||      | |  || | |
Db        361 TTTGACTACTATAAATGCTGAAAAAAGTTTTTGTGAAAAATTCCTACCTATGCAAGGTTT 420

Qy        415 AGAACAGATGATGAACTTTATGTGTTTAATAATA 448
              |||| |||   | ||  | |  ||   ||  | |
Db        421 AGAAAAGAAATTAAAGGTGAATTGAAAAAACAAA 454


RESULT 6
US-09-960-870-17
; Sequence 17, Application US/09960870
; Publication No. US20030134281A1
; GENERAL INFORMATION:
;  APPLICANT: Evans, Glen
;  TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
;  TITLE OF INVENTION:  USE
;  FILE REFERENCE: P-EA 4738
;  CURRENT APPLICATION NUMBER: US/09/960,870
;  CURRENT FILING DATE:  2001-09-20
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 2601
;   TYPE: DNA
;   ORGANISM: M. genitalium
US-09-960-870-17

  Query Match             5.0%;  Score 49.6;  DB 3;  Length 2601;
  Best Local Similarity   46.0%;  Pred. No. 3.4;
  Matches  209;  Conservative    0;  Mismatches  239;  Indels    6;  Gaps    1;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||    ||||  |   ||| || ||||| ||| |   |  || | ||||| |   |
Db        519 ATGAAATTATCTGTAATTATACCTACTTACAATTGTGCATCATTTATTGAAAAAGCAATT 578

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              |||   ||   |||  |   |    | |||| || || ||  | |  ||    ||| ||
Db        579 AATTCAATTGTTAAAAATAGACCTAATGATTTGGAAATAGAAGTTTTAATTATTGATGAT 638

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              |  || | |||   |||  |  || | ||| |          || |   ||||     |
Db        639 GGATCAATTGACAATACTAACAAAGTTATTAAGAAAATTCAAGACCAAATTAATAATTTA 698

Qy        181 ATAATATATCGGCGT------AATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCT 234
              |   |  |      |      | ||  | ||||   ||     | ||  |||| | ||
Db        699 ACTTTGCAGTATTTTTACAAAAGTAATGGTAACTGGGGTAGTGTTATTAATTATGTTAGA 758

Qy        235 TCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCG 294
                || || || ||||| |          |||| |  |||||||    | || |
Db        759 AACAATAAACTAGCAAAAGGGGAATGAGTAACAGTATTGGATAGTGATGACATTTTTTCA 818

Qy        295 AAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGT 354
              ||| |  |  |  | || || ||||  |||  | |    |||  | ||       |   |
Db        819 AAAAAAACAATTTCTATTTTTCAAAAATATGCCCAAAAACAAAGATATGATGCGATTATT 878

Qy        355 GACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTC 414
                | | |   | ||  |      ||  |   | ||  |||      | |  || | |
Db        879 TTTGACTACTATAAATGCTGAAAAAAGTTTTTGTGAAAAATTCCTACCTATGCAAGGTTT 938

Qy        415 AGAACAGATGATGAACTTTATGTGTTTAATAATA 448
              |||| |||   | ||  | |  ||   ||  | |
Db        939 AGAAAAGAAATTAAAGGTGAATTGAAAAAACAAA 972


RESULT 7
US-09-960-858-17
; Sequence 17, Application US/09960858
; Publication No. US20030138777A1
; GENERAL INFORMATION:
;  APPLICANT: Evans, Glen
;  TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
;  TITLE OF INVENTION:  USE
;  FILE REFERENCE: P-EA 4974
;  CURRENT APPLICATION NUMBER: US/09/960,858
;  CURRENT FILING DATE:  2001-09-20
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 2601
;   TYPE: DNA
;   ORGANISM: M. genitalium
US-09-960-858-17

  Query Match             5.0%;  Score 49.6;  DB 3;  Length 2601;
  Best Local Similarity   46.0%;  Pred. No. 3.4;
  Matches  209;  Conservative    0;  Mismatches  239;  Indels    6;  Gaps    1;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||    ||||  |   ||| || ||||| ||| |   |  || | ||||| |   |
Db        519 ATGAAATTATCTGTAATTATACCTACTTACAATTGTGCATCATTTATTGAAAAAGCAATT 578

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              |||   ||   |||  |   |    | |||| || || ||  | |  ||    ||| ||
Db        579 AATTCAATTGTTAAAAATAGACCTAATGATTTGGAAATAGAAGTTTTAATTATTGATGAT 638

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              |  || | |||   |||  |  || | ||| |          || |   ||||     |
Db        639 GGATCAATTGACAATACTAACAAAGTTATTAAGAAAATTCAAGACCAAATTAATAATTTA 698

Qy        181 ATAATATATCGGCGT------AATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCT 234
              |   |  |      |      | ||  | ||||   ||     | ||  |||| | ||
Db        699 ACTTTGCAGTATTTTTACAAAAGTAATGGTAACTGGGGTAGTGTTATTAATTATGTTAGA 758

Qy        235 TCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCG 294
                || || || ||||| |          |||| |  |||||||    | || |
Db        759 AACAATAAACTAGCAAAAGGGGAATGAGTAACAGTATTGGATAGTGATGACATTTTTTCA 818

Qy        295 AAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGT 354
              ||| |  |  |  | || || ||||  |||  | |    |||  | ||       |   |
Db        819 AAAAAAACAATTTCTATTTTTCAAAAATATGCCCAAAAACAAAGATATGATGCGATTATT 878

Qy        355 GACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTC 414
                | | |   | ||  |      ||  |   | ||  |||      | |  || | |
Db        879 TTTGACTACTATAAATGCTGAAAAAAGTTTTTGTGAAAAATTCCTACCTATGCAAGGTTT 938

Qy        415 AGAACAGATGATGAACTTTATGTGTTTAATAATA 448
              |||| |||   | ||  | |  ||   ||  | |
Db        939 AGAAAAGAAATTAAAGGTGAATTGAAAAAACAAA 972


RESULT 8
US-10-251-668-17
; Sequence 17, Application US/10251668
; Publication No. US20040063097A1
; GENERAL INFORMATION:
;  APPLICANT: Evans, Glen
;  TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
;  TITLE OF INVENTION:  USE
;  FILE REFERENCE: P-EA 5441
;  CURRENT APPLICATION NUMBER: US/10/251,668
;  CURRENT FILING DATE:  2002-09-20
;  PRIOR APPLICATION NUMBER: US 09/960,607
;  PRIOR FILING DATE: 2001-09-20
```

```
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 2601
;   TYPE: DNA
;   ORGANISM: M. genitalium
US-10-251-668-17

  Query Match             5.0%;  Score 49.6;  DB 7;  Length 2601;
  Best Local Similarity   46.0%;  Pred. No. 3.4;
  Matches  209;  Conservative    0;  Mismatches  239;  Indels    6;  Gaps    1;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||     ||||  |   ||| || ||||| ||| |    |  || | ||||| |     |
Db        519 ATGAAATTATCTGTAATTATACCTACTTACAATTGTGCATCATTTATTGAAAAAGCAATT 578

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              |||    ||    |||  |    |      | |||| || || ||  | |  ||    ||| ||
Db        579 AATTCAATTGTTAAAAATAGACCTAATGATTTGGAAATAGAAGTTTTAATTATTGATGAT 638

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              |  || | |||   |||  |  || | ||| |                ||  |   ||||    |
Db        639 GGATCAATTGACAATACTAACAAAGTTATTAAGAAAATTCAAGACCAAATTAATAATTTA 698

Qy        181 ATAATATATCGGCGT------AATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCT 234
              |   | |      |      | ||  | ||||   ||     | ||  |||| | ||
Db        699 ACTTTGCAGTATTTTTACAAAAGTAATGGTAACTGGGGTAGTGTTATTAATTATGTTAGA 758

Qy        235 TCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCG 294
                  || || || ||||| |          |||| |  |||||||   | || |
Db        759 AACAATAAACTAGCAAAAGGGGAATGAGTAACAGTATTGGATAGTGATGACATTTTTTCA 818

Qy        295 AAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGT 354
              ||| |  |  |  | || || ||||  |||  | |    |||  | ||        |  |
Db        819 AAAAAAACAATTTCTATTTTTCAAAAATATGCCCAAAAACAAAGATATGATGCGATTATT 878

Qy        355 GACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTC 414
                |  | |   | ||  |      ||  |   | ||  |||     | |  || |
Db        879 TTTGACTACTATAAATGCTGAAAAAAGTTTTTGTGAAAAATTCCTACCTATGCAAGGTTT 938

Qy        415 AGAACAGATGATGAACTTTATGTGTTTAATAATA 448
              |||| |||   | ||  | |  ||   ||  | |
Db        939 AGAAAAGAAATTAAAGGTGAATTGAAAAAACAAA 972


RESULT 9
US-10-205-220-1
; Sequence 1, Application US/10205220
; Publication No. US20030170663A1
; GENERAL INFORMATION:
;  APPLICANT: Fraser et al.
;  TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
;  TITLE OF INVENTION:  Thereof, and Uses Thereof
;  FILE REFERENCE: PB193P1D1
;  CURRENT APPLICATION NUMBER: US/10/205,220
;  CURRENT FILING DATE:  2002-07-26
;  PRIOR APPLICATION NUMBER: US 08/545,528
;  PRIOR FILING DATE: 1995-10-19
;  PRIOR APPLICATION NUMBER: US 08/488,018
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: US 08/473,545
;  PRIOR FILING DATE: 1995-06-07
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 580073
;   TYPE: DNA
;   ORGANISM: Mycoplasma genitalium
US-10-205-220-1

  Query Match             5.0%;  Score 49.6;  DB 6;  Length 580073;
  Best Local Similarity   46.0%;  Pred. No. 28;
  Matches  209;  Conservative    0;  Mismatches  239;  Indels    6;  Gaps    1;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||     ||||  |   ||| || ||||| ||| |    |  || | ||||| |     |
Db      67651 ATGAAATTATCTGTAATTATACCTACTTACAATTGTGCATCATTTATTGAAAAAGCAATT 67710

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              |||    ||    |||  |    |      | |||| || || ||  | |  ||    ||| ||
Db      67711 AATTCAATTGTTAAAAATAGACCTAATGATTTGGAAATAGAAGTTTTAATTATTGATGAT 67770

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              |  || | |||   |||  |  || | ||| |                ||  |   ||||    |
Db      67771 GGATCAATTGACAATACTAACAAAGTTATTAAGAAAATTCAAGACCAAATTAATAATTTA 67830

Qy        181 ATAATATATCGGCGT------AATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCT 234
              |   | |      |      | ||  | ||||   ||     | ||  |||| | ||
Db      67831 ACTTTGCAGTATTTTTACAAAAGTAATGGTAACTGGGGTAGTGTTATTAATTATGTTAGA 67890

Qy        235 TCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCG 294
                  || || || ||||| |          |||| |  |||||||   | || |
Db      67891 AACAATAAACTAGCAAAAGGGGAATGAGTAACAGTATTGGATAGTGATGACATTTTTTCA 67950

Qy        295 AAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGT 354
              ||| |  |  |  | || || ||||  |||  | |    |||  | ||        |  |
Db      67951 AAAAAAACAATTTCTATTTTTCAAAAATATGCCCAAAAACAAAGATATGATGCGATTATT 68010

Qy        355 GACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTC 414
                |  | |   | ||  |      ||  |   | ||  |||     | |  || |
Db      68011 TTTGACTACTATAAATGCTGAAAAAAGTTTTTGTGAAAAATTCCTACCTATGCAAGGTTT 68070

Qy        415 AGAACAGATGATGAACTTTATGTGTTTAATAATA 448
              |||| |||   | ||  | |  ||   ||  | |
Db      68071 AGAAAAGAAATTAAAGGTGAATTGAAAAAACAAA 68104


RESULT 10
US-09-070-927A-443/c
; Sequence 443, Application US/09070927A
; Patent No. US20020120116A1
;   GENERAL INFORMATION:
;        APPLICANT: Charles A. Kunsch
;                   Patrick J. Dillon
;                   Steven Barash
;        TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
;        NUMBER OF SEQUENCES: 982
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/070,927A
;             FILING DATE: 04-May-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/046,655
;             FILING DATE: 1997-05-16
;             APPLICATION NUMBER: 60/044,031
;             FILING DATE: 1997-05-06
;             APPLICATION NUMBER: 60/066,009
;             FILING DATE: 1997-11-14
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Kenley K. Hoover
;             REGISTRATION NUMBER: 40,302
;             REFERENCE/DOCKET NUMBER: PB369
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (301) 309-8504
```

```
;             TELEFAX: (301) 309-8512
;   INFORMATION FOR SEQ ID NO: 443:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2012 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 443:
US-09-070-927A-443

  Query Match             4.9%;  Score 49.2;  DB 3;  Length 2012;
  Best Local Similarity   64.1%;  Pred. No. 3.7;
  Matches   91;  Conservative    1;  Mismatches   44;  Indels    6;  Gaps    1;

Qy          5 TTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATA 64
              | || ||| |||||||| | || |||||| || || ||||||| |  ||    ||  |||
Db       1890 TGATCTCAGTTTGTATTGCGACATATAATGGAGAAAAATATCTCGCGGAACAATTAGATA 1831

Qy         65 GTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCT 124
              ||||   |   ||:|      | | |  |||   |||  ||  ||| |||||| |||  |
Db       1830 GTATTCTTTTACARG------TCAGTGAAGARGATGAACTAATTATTTCAGATGATGGTT 1777

Qy        125 CTACTGATGGTACAGAGGAAAT 146
              ||||||||  |||   ||||||
Db       1776 CTACTGATCATACGTTGGAAAT 1755


RESULT 11
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             4.7%;  Score 47.4;  DB 3;  Length 640681;
  Best Local Similarity   51.7%;  Pred. No. 86;
  Matches  108;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        703 AGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAA 762
              | |  ||||  ||||  | |||| |  | | |  || | |  |  ||     | ||
Db     325630 AAAATAATTACAATTCTATCTATTATAGGTATTAGTATAACGACAGCTTCACTAATTATT 325689

Qy        763 AATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATAT 822
              | || |||| ||| |   | || ||| ||   |||  |||| | |    ||  ||   |
Db     325690 ATTACATCTATAATAAATGGATCTGAGAAAAATTTTGAAAAAAATATTTTATCATTTATT 325749

Qy        823 ACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTT 882
              ||  |||    ||  |   | |       |  ||  | |||||    | ||  |||||  |
Db     325750 CCACATTTAATAATAACTAATAAAAATCAATGTATTAAAAAAGATCAATTTCCTGAAAAT 325809

Qy        883 TATGCAAAGCTAGGTTGTAATAAAAATAT 911
                |  |||  ||  |  || |||||||||
Db     325810 ATTTTAAAATTAAATAATATTAAAAATAT 325838


RESULT 12
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (379615)
US-10-312-841-2

  Query Match             4.7%;  Score 47.2;  DB 6;  Length 3673778;
  Best Local Similarity   47.9%;  Pred. No. 1.8e+02;
  Matches  136;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        631 TTGCATTATATATCAAAACCACTCGTAATATGCCGAGGAGATAATGATAGTTTCGAGAAG 690
              ||| ||| | | | | ||  | |   || ||         || | ||   | ||   | |
Db    3011564 TTGAATTTTTTGTTATAATGAATAAGAAAATAAAATAAAGTTTATTTGAATTGAAAAGAA 3011623

Qy        691 AAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGAT 750
              | | | ||    |      || ||  |||| | |||  |||| || || |    ||  ||
Db    3011624 AGAAGTAAAATTATTTTCGTTATAGATGATATGATTTTATATATAGAAATTTAAAAGAAT 3011683

Qy        751 TTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGA 810
              || |    |||| |||     | |||  |   | | | | || |  | || || |   |
Db    3011684 TTATTTTAAAAAAGTATTATAATTAAAGAATTTTATTATATTGGTAGGATATAAAATTAA 3011743

Qy        811 CCATGGTTATATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGAT 870
                  |  ||     | |   |||     | |||  | |  |   |||||||| || |
Db    3011744 AATATAAAAATTAATTGTGTTTTTATATAGTATTAATGAATGAAAAATGAAAAATGAAAA 3011803

Qy        871 TTATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATATGAT 914
              || |   | | |     |||  ||| ||  |||  |   | |||
Db    3011804 TTTTGAAATTATATATAAAAAGTAGTTTAAAATGTATTAAAGAT 3011847


RESULT 13
US-10-433-793-46
; Sequence 46, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/433,793
;  CURRENT FILING DATE:  2003-06-06
;  NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 46
;   LENGTH: 10048
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-46

  Query Match             4.7%;  Score 47;  DB 7;  Length 10048;
  Best Local Similarity   49.8%;  Pred. No. 20;
  Matches  147;  Conservative    0;  Mismatches  145;  Indels    3;  Gaps    1;

Qy        711 TTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATC 770
              |||| || ||||  ||  |||  | |  |||      |  || |    |||| ||||
Db        663 TTTATTTAATTTATTTTAATAAATGATTTTATAATGTATTTATGTATTTAAAAGTATAGA 722

Qy        771 TTTAAAA---CGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAAC 827
```

```
Db        723 GATATAAGTTTTTATATATGAAGTTGTTATGGTTATAGAAAGAAAATAAATATTTGTGTA 782

Qy        828 TTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGC 887

Db        783 TATAGAGATAGTATGAAATTATTTTATAGTGAAAATAATAGTTTTTGGAAAAAGTTTTGA 842

Qy        888 AAAGCTAGGTTGTAATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATA 947

Db        843 AAATTAGAGATGGTGTTATTATTAGATAATATAAGTGTTGGGGTATAGGGTAGTTTTTTA 902

Qy        948 TGCAGTGAAAAATATTACCGTGCTTAAGAATTTTACTAAACGGATAATTAAGTAG 1002

Db        903 TGTATTTTATAGTATTTGTATGGTTTGTGAGTTTTATTTAATGATATAAAATGAG 957


RESULT 14
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3294164)
US-10-312-841-1

  Query Match             4.7%;  Score 47;  DB 6;  Length 3673778;
  Best Local Similarity   45.2%;  Pred. No. 2e+02;
  Matches  173;  Conservative    0;  Mismatches  210;  Indels    0;  Gaps    0;

Qy        426 TGAACTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATC 485

Db    3192006 TTAATACTATATAAAAAACACTACTATATTCAAAAAAACGTATAATAATCTAATAATATA 3191947

Qy        486 TATTGGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGGGATGC 545

Db    3191946 TATTTCTATATTAATACAATACAATATTTCTTTACTAATACATTAAAAAACATAATTTTT 3191887

Qy        546 CATTGATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAG 605

Db    3191886 TAACAAATTTAACAACTCCCATAATTATAAATAATATTTTAAAATATCTATAACAACTAT 3191827

Qy        606 CGTATTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATATGCCG 665

Db    3191826 AATATAATATAAAAATATTAATTAATAATATTTTCTATTAATAACAATATTACAACTACT 3191767

Qy        666 AGGAGATAATGATAGTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTAT 725

Db    3191766 AACTAATACTACAAAAATTTATTTTCTACATTCATAAAATTAAATTTCAATAAAAATTTT 3191707

Qy        726 TGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATT 785

Db    3191706 TTAAAAATAAAATTAACTTTTTTTTATTCCATCCAAATTTATAAACTTCTAAATTATTTT 3191647

Qy        786 TGAAAATGTTTTGCTAAAAGAGA 808

Db    3191646 TTTTTATTTTTTATTAATACATA 3191624


RESULT 15
US-10-147-603-204/c
; Sequence 204, Application US/10147603
; Publication No. US20030219750A1
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE: 53.US16.DIV
;  CURRENT APPLICATION NUMBER: US/10/147,603
;  CURRENT FILING DATE:  2000-03-30
;  PRIOR APPLICATION NUMBER: 09/539,333
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: US 60/126,903
;  PRIOR FILING DATE: 1999-03-30
;  PRIOR APPLICATION NUMBER: US 60/131,971
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/132,065
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/143,928
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 60/145,915
;  PRIOR FILING DATE: 1999-07-27
;  PRIOR APPLICATION NUMBER: US 60/146,453
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/146,452
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/162,288
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: US 09/416,384
;  PRIOR FILING DATE: 1999-10-12
;  NUMBER OF SEQ ID NOS: 234
;  SOFTWARE: Patent.pm
; SEQ ID NO 204
;   LENGTH: 3001
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 1501
;   OTHER INFORMATION: 99-26267-524 : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 1481..1500
;   OTHER INFORMATION: 99-26267-524.mis1,
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 1502..1521
;   OTHER INFORMATION: 99-26267-524.mis2,  complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 983..1002
;   OTHER INFORMATION: upstream amplification primer
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1553..1573
;   OTHER INFORMATION: downstream amplification primer, complement
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 1489..1513
;   OTHER INFORMATION: 99-26267-524  probe
US-10-147-603-204

  Query Match             4.7%;  Score 46.8;  DB 6;  Length 3001;
  Best Local Similarity   49.2%;  Pred. No. 14;
  Matches  123;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy        710 TTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATAT 769

Db       2198 TATATATGAATATAATGTATAATGTAAAATTATATACATTATTAAACTATAAAATTATAT 2139
```

```
Qy        770 CTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTT 829
Db       2138 AGTAAAATTATATAATGTATATTTAATATTTTATTTATATATAATGTATAAATATACATA 2079

Qy        830 TGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAA 889
Db       2078 TTTATATGTAATGTATATATATTATTTATACACATATATACACATATATAATATATACAT 2019

Qy        890 AGCTAGGTTGTAATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATG 949
Db       2018 ATATTCCATTTTATATATATACACACAAATATGATCTGGGGAAAAGGTTTCGTGCATACA 1959

Qy        950 CAGTGAAAAA 959
Db       1958 CACACACAAA 1949
```

Search completed: May 12, 2006, 05:35:25
Job time : 655.566 secs

GenCore version 5.1.8
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2006, 03:52:36 ; Search time 510.518 Seconds
(without alignments)
8004.956 Million cell updates/sec

Title: US-09-423-093B-4_COPY_14059_15060
Perfect score: 1002
Sequence: 1 atgcttatatcattttgtat.....ctaaacggataattaagtag 1002

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9312410 seqs, 2039259788 residues

Total number of hits satisfying chosen parameters: 18624820

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA_New:*
1: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq1:*
2: /SIDS5/ptodata/2/pubpna/US06_NEW_PUB.seq:*
3: /SIDS5/ptodata/2/pubpna/US07_NEW_PUB.seq:*
4: /SIDS5/ptodata/2/pubpna/US08_NEW_PUB.seq:*
5: /SIDS5/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
6: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq:*
7: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq1:*
8: /SIDS5/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
9: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq:*
10: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq1:*
11: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
12: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq3:*
13: /SIDS5/ptodata/2/pubpna/US10_NEW_PUB.seq4:*
14: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq:*
15: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq2:*
16: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq3:*
17: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq4:*
18: /SIDS5/ptodata/2/pubpna/US11_NEW_PUB.seq5:*
19: /SIDS5/ptodata/2/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	51.6	5.1	592	7	US-09-925-065A-651832	Sequence 651832,
2	49.6	5.0	2601	18	US-11-143-564-17	Sequence 17, Appl
c 3	46.8	4.7	3001	17	US-11-145-703-204	Sequence 204, App
4	45.8	4.6	395	12	US-10-301-480-247683	Sequence 247683,
5	45.8	4.6	395	12	US-10-301-480-861092	Sequence 861092,
6	45.8	4.6	415	7	US-09-925-065A-154135	Sequence 154135,
c 7	45.4	4.5	604	7	US-09-925-065A-226132	Sequence 226132,
8	45.4	4.5	610	12	US-10-301-480-310628	Sequence 310628,
9	45.4	4.5	610	12	US-10-301-480-924037	Sequence 924037,
c 10	43.2	4.3	721	7	US-09-925-065A-63645	Sequence 63645, A
c 11	43.2	4.3	721	11	US-10-301-480-164883	Sequence 164883,
c 12	43.2	4.3	721	12	US-10-301-480-778292	Sequence 778292,
13	43	4.3	930	18	US-11-079-463-1998	Sequence 1998, Ap
c 14	42.8	4.3	49979	10	US-10-995-561-13443	Sequence 13443, A
c 15	42.4	4.2	519	12	US-10-301-480-564593	Sequence 564593,
c 16	42.4	4.2	519	12	US-10-301-480-1178002	Sequence 1178002,
17	42.2	4.2	633	7	US-09-925-065A-99890	Sequence 99890, A
18	42.2	4.2	635	11	US-10-301-480-65390	Sequence 65390, A
19	42.2	4.2	635	11	US-10-301-480-200553	Sequence 200553,
20	42.2	4.2	635	12	US-10-301-480-678799	Sequence 678799,
21	42.2	4.2	635	12	US-10-301-480-813962	Sequence 813962,
c 22	42.2	4.2	879	12	US-10-301-480-572204	Sequence 572204,
c 23	42.2	4.2	879	12	US-10-301-480-1185613	Sequence 1185613,
24	42.2	4.2	6866	10	US-10-240-708-20	Sequence 20, Appl
c 25	42	4.2	987	12	US-10-301-480-545289	Sequence 545289,
c 26	42	4.2	987	12	US-10-301-480-1158698	Sequence 1158698,
27	42	4.2	6499	17	US-11-011-332A-90	Sequence 90, Appl
28	41.6	4.2	529	7	US-09-925-065A-666232	Sequence 666232,
c 29	41.4	4.1	2362	7	US-09-925-065A-673953	Sequence 673953,
30	41.4	4.1	8093	10	US-10-240-708-31	Sequence 31, Appl
31	41.2	4.1	62444	11	US-10-330-773-551	Sequence 551, App
32	41.2	4.1	173602	17	US-11-121-086-25	Sequence 25, Appl
33	41	4.1	554	7	US-09-925-065A-462096	Sequence 462096,
c 34	41	4.1	563	12	US-10-301-480-518836	Sequence 518836,
c 35	41	4.1	563	12	US-10-301-480-1132245	Sequence 1132245,
36	41	4.1	1535	7	US-09-925-065A-60008	Sequence 60008, A
37	41	4.1	1535	11	US-10-301-480-161246	Sequence 161246,
38	41	4.1	1535	12	US-10-301-480-774655	Sequence 774655,
39	41	4.1	139054	17	US-11-121-086-96	Sequence 96, Appl
40	40.8	4.1	966	18	US-11-079-463-1999	Sequence 1999, Ap
c 41	40.8	4.1	1542	10	US-10-750-185-29103	Sequence 29103, A
c 42	40.8	4.1	1542	10	US-10-750-623-29103	Sequence 29103, A
43	40.6	4.1	643	7	US-09-925-065A-756015	Sequence 756015,
44	40.6	4.1	28536	17	US-11-011-332A-151	Sequence 151, App
c 45	40.4	4.0	201	10	US-10-995-561-52720	Sequence 52720, A

ALIGNMENTS

RESULT 1
US-09-925-065A-651832/c
; Sequence 651832, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 651832
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-651832

Query Match 5.1%; Score 51.6; DB 7; Length 592;
Best Local Similarity 53.5%; Pred. No. 2;
Matches 108; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```
Qy        708 AATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATAT 767
              || ||| ||| | ||||    || | ||   ||| |  || ||||     |||||
Db        496 AAGTTTTATTTACATTATATTTTAGTCTAAGCGAGCAATAGAATTTTGTCTAAAAAAAAC 437
```

```
Qy        768 ATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAAC 827
              |    |  |||   ||||| |||||  |||  | | | | | |  ||||| ||   || |
Db        436 AATGTATACACCTTCATTTAAAAATACTTTCTTGATAAATAAAAAATGGTAATGATCATC 377

Qy        828 TTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGC 887
              | | || | ||  || ||      | | | |  | || |||||  ||   | ||
Db        376 TGAGCCTTTAGCTAGTCATAATCTTTTTTTTTATTAGAGGGATTTGCCTTTCTGTTGATG 317

Qy        888 AAAGCTAGGTTGTAATAAAAAT 909
              |  |||   |||  | ||  ||
Db        316 ACTGCTGATTTGCCAGAATGAT 295


RESULT 2
US-11-143-564-17
; Sequence 17, Application US/11143564
; Publication No. US20060078907A1
; GENERAL INFORMATION:
;  APPLICANT: Evans, Glen
;  TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
;  TITLE OF INVENTION:  USE
;  FILE REFERENCE: P-EA 5441
;  CURRENT APPLICATION NUMBER: US/11/143,564
;  CURRENT FILING DATE:  2005-06-02
;  PRIOR APPLICATION NUMBER: US/10/251,668
;  PRIOR FILING DATE: 2002-09-20
;  PRIOR APPLICATION NUMBER: US 09/960,607
;  PRIOR FILING DATE: 2001-09-20
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 2601
;   TYPE: DNA
;   ORGANISM: M. genitalium
US-11-143-564-17

  Query Match             5.0%;  Score 49.6;  DB 18;  Length 2601;
  Best Local Similarity   46.0%;  Pred. No. 6.4;
  Matches  209;  Conservative    0;  Mismatches  239;  Indels    6;  Gaps    1;

Qy          1 ATGCTTATATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTG 60
              |||    ||||  |   ||| || ||||| ||| |   |  || | |||| | |     |
Db        519 ATGAAATTATCTGTAATTATACCTACTTACAATTGTGCATCATTTATTGAAAAAGCAATT 578

Qy         61 AATAGTATAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAAT 120
              |||   ||  |||  |  |     | |||| || || ||  | |  ||    ||| ||
Db        579 AATTCAATTGTTAAAAATAGACCTAATGATTTGGAAATAGAAGTTTTAATTATTGATGAT 638

Qy        121 GCCTCTACTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCA 180
              |  || | |||   |||  |  || | ||| |           ||| |   ||||     |
Db        639 GGATCAATTGACAATACTAACAAAGTTATTAAGAAAATTCAAGACCAAATTAATAATTTA 698

Qy        181 ATAATATATCGGCGT------AATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCT 234
              |   |  |       |      | ||  | ||||   ||     | ||  |||| | ||
Db        699 ACTTTGCAGTATTTTTACAAAAGTAATGGTAACTGGGGTAGTGTTATTAATTATGTTAGA 758

Qy        235 TCAGTATCCCTTGCGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCG 294
                  || || || ||||| |          |||| |  |||||||   | || |
Db        759 AACAATAAACTAGCAAAAGGGGAATGAGTAACAGTATTGGATAGTGATGACATTTTTTCA 818

Qy        295 AAAGACTCGTTAGCGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGT 354
              ||| |  |  | | || || |||| ||| | |    ||| | ||      |   |
Db        819 AAAAAAACAATTTCTATTTTTCAAAAATATGCCCAAAAACAAAGATATGATGCGATTATT 878

Qy        355 GACAGAAAAGAGACCGGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTC 414
                |  | |   | || |      ||  |   | ||  |||      | |  || |
Db        879 TTTGACTACTATAAATGCTGAAAAAAGTTTTTGTGAAAAATTCCTACCTATGCAAGGTTT 938

Qy        415 AGAACAGATGATGAACTTTATGTGTTTAATAATA 448
              |||| |||   | ||  | |  ||   || | |
Db        939 AGAAAAGAAATTAAAGGTGAATTGAAAAAACAAA 972
```

```
RESULT 3
US-11-145-703-204/c
; Sequence 204, Application US/11145703
; Publication No. US20050260667A1
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE: 53.US16.DIV
;  CURRENT APPLICATION NUMBER: US/11/145,703
;  CURRENT FILING DATE:  2000-03-30
;  PRIOR APPLICATION NUMBER: US/10/147,603
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: 09/539,333
;  PRIOR FILING DATE: 2000-03-30
;  PRIOR APPLICATION NUMBER: US 60/126,903
;  PRIOR FILING DATE: 1999-03-30
;  PRIOR APPLICATION NUMBER: US 60/131,971
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/132,065
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/143,928
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 60/145,915
;  PRIOR FILING DATE: 1999-07-27
;  PRIOR APPLICATION NUMBER: US 60/146,453
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/146,452
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/162,288
;  PRIOR FILING DATE: 1999-10-28
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 234
;  SOFTWARE: Patent.pm
; SEQ ID NO 204
;   LENGTH: 3001
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 1501
;   OTHER INFORMATION: 99-26267-524 : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 1481..1500
;   OTHER INFORMATION: 99-26267-524.mis1,
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 1502..1521
;   OTHER INFORMATION: 99-26267-524.mis2,  complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 983..1002
;   OTHER INFORMATION: upstream amplification primer
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1553..1573
;   OTHER INFORMATION: downstream amplification primer, complement
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 1489..1513
;   OTHER INFORMATION: 99-26267-524  probe
US-11-145-703-204

  Query Match             4.7%;  Score 46.8;  DB 17;  Length 3001;
  Best Local Similarity   49.2%;  Pred. No. 22;
  Matches  123;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy        710 TTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATAT 769
```

```
Db       2198 TATATATGAATATAATGTATAATGTAAAATTATATACATTATTAAACTATAAAATTATAT 2139

Qy        770 CTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACTT 829

Db       2138 AGTAAAATTATATAATGTATATTTAATATTTTATTTATATATAATGTATAAATATACATA 2079

Qy        830 TGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTATGCAA 889

Db       2078 TTTATATGTAATGTATATATATTATTTATACACATATATACACATATATAATATATACAT 2019

Qy        890 AGCTAGGTTGTAATAAAAATATGATCAACACTGTACTTCGATTTGGGAAACTAGCATATG 949

Db       2018 ATATTCCATTTTATATATATACACACAAATATGATCTGGGGAAAAGGTTTCGTGCATACA 1959

Qy        950 CAGTGAAAAA 959

Db       1958 CACACACAAA 1949
```

RESULT 4
US-10-301-480-247683
; Sequence 247683, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 247683
; LENGTH: 395
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-247683

Query Match 4.6%; Score 45.8; DB 12; Length 395;
Best Local Similarity 50.6%; Pred. No. 23;
Matches 137; Conservative 0; Mismatches 132; Indels 2; Gaps 1;

```
Qy        636 TTATATATCAAAACCACTCGTAATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGG 695

Db         13 TTGTATATCTGAACCTGTCTAAATGTAAAAAGGTACAGTAAAAATATTATAGTATACTAT 72

Qy        696 AAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTA 755

Db         73 TATATTAATGTTAATAGTATATTATATTATTATTCATAATATTTTTATTAATAAATACAA 132

Qy        756 CAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATG 815

Db        133 TATTTYATTAATAAATTAATAAATATAATAATAAAATATATTACTATAATATATTTTAAT 192

Qy        816 GTTATATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATC 875

Db        193 GTATAATATA--TTATAATATGTATTATAATAAATATAATAATAAAATATGATATTAATA 250

Qy        876 TGAATTTTATGCAAAGCTAGGTTGTAATAAA 906

Db        251 TTAATATTATACTATAATATTATATATTATA 281
```

RESULT 5
US-10-301-480-861092
; Sequence 861092, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 861092
; LENGTH: 395
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-861092

Query Match 4.6%; Score 45.8; DB 12; Length 395;
Best Local Similarity 50.6%; Pred. No. 23;
Matches 137; Conservative 0; Mismatches 132; Indels 2; Gaps 1;

```
Qy        636 TTATATATCAAAACCACTCGTAATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGG 695

Db         13 TTGTATATCTGAACCTGTCTAAATGTAAAAAGGTACAGTAAAAATATTATAGTATACTAT 72

Qy        696 AAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTA 755

Db         73 TATATTAATGTTAATAGTATATTATATTATTATTCATAATATTTTTATTAATAAATACAA 132

Qy        756 CAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATG 815

Db        133 TATTTYATTAATAAATTAATAAATATAATAATAAAATATATTACTATAATATATTTTAAT 192

Qy        816 GTTATATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATC 875

Db        193 GTATAATATA--TTATAATATGTATTATAATAAATATAATAATAAAATATGATATTAATA 250

Qy        876 TGAATTTTATGCAAAGCTAGGTTGTAATAAA 906

Db        251 TTAATATTATACTATAATATTATATATTATA 281
```

RESULT 6
US-09-925-065A-154135
; Sequence 154135, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 154135
; LENGTH: 415
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-154135

Query Match 4.6%; Score 45.8; DB 7; Length 415;

Best Local Similarity 50.6%; Pred. No. 23;
Matches 137; Conservative 0; Mismatches 132; Indels 2; Gaps 1;

```
Qy        636 TTATATATCAAAACCACTCGTAATATGCCGAGGAGATAATGATAGTTTCGAGAAGAAAGG 695
              || ||||||  ||||  ||  |||  |     ||         | |  ||  || | |
Db         13 TTGTATATCTGAACCTGTCTAAATGTAAAAAGGTACAGTAAAAATATTATAGTATACTAT 72

Qy        696 AAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTA 755
              |      |    |||  ||  | || |||||    ||  |  ||  |||| | |   |
Db         73 TATATTAATGTTAATAGTATATTATATTATTATTCATAATATTTTTATTAATAAATACAA 132

Qy        756 CAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATG 815
              | |  |  |||  || | ||  |  ||   ||||| | || ||| || | |    |
Db        133 TATTTYATTAATAAATTAATAAATATAATAATAAAATATATTACTATAATATATTTTAAT 192

Qy        816 GTTATATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATC 875
              ||   ||| |  ||    ||||   | | ||    ||| | || ||| | || ||| ||
Db        193 GTATAATATA--TTATAATATGTATTATAATAAATATAATAATAAAATATGATATTAATA 250

Qy        876 TGAATTTTATGCAAAGCTAGGTTGTAATAAA 906
              | ||| |||| | |   ||   | || || |
Db        251 TTAATATTATACTATAATATTATATATTATA 281
```

RESULT 7
US-09-925-065A-226132/c
; Sequence 226132, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 226132
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-226132

Query Match 4.5%; Score 45.4; DB 7; Length 604;
Best Local Similarity 53.1%; Pred. No. 29;
Matches 94; Conservative 1; Mismatches 82; Indels 0; Gaps 0;

```
Qy        709 ATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATA 768
              ||||||  |  |||  |||    ||    ||  |||||   ||   |||||    ||  |  ||
Db        422 ATTTTATTAAATTATATATAATTCATATAATTACATATGAATTTTACTAGGGATAGGTTA 363

Qy        769 TCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACT 828
              | ||||  ||  | |  ||   ||    |||  || | |   |   ||    | |  |
Db        362 GTATGAAAAACAGTTTATAGAAGGTTTACATTAATTGAAAAATAACCTTTTGTCATACTT 303

Qy        829 TTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTAT 885
              |||  |||  |:  |||       |||  ||  ||||   ||||||  |||||  || |
Db        302 TTGTTTATTTCRCCTTACAAGCTTAGGCATTTAAAACTTGATTTACCTGAAAGTTGT 246
```

RESULT 8
US-10-301-480-310628
; Sequence 310628, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 310628
; LENGTH: 610
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-310628

Query Match 4.5%; Score 45.4; DB 12; Length 610;
Best Local Similarity 53.1%; Pred. No. 29;
Matches 94; Conservative 1; Mismatches 82; Indels 0; Gaps 0;

```
Qy        709 ATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATA 768
              ||||||  |  |||  |||    ||    ||  |||||   ||   |||||    ||  |  ||
Db        189 ATTTTATTAAATTATATATAATTCATATAATTACATATGAATTTTACTAGGGATAGGTTA 248

Qy        769 TCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACT 828
              | ||||  ||  | |  ||   ||    |||  || | |   |   ||    | |  |
Db        249 GTATGAAAAACAGTTTATAGAAGGTTTACATTAATTGAAAAATAACCTTTTGTCATACTT 308

Qy        829 TTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTAT 885
              |||  |||  |:  |||       |||  ||  ||||   ||||||  |||||  || |
Db        309 TTGTTTATTTCRCCTTACAAGCTTAGGCATTTAAAACTTGATTTACCTGAAAGTTGT 365
```

RESULT 9
US-10-301-480-924037
; Sequence 924037, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 924037
; LENGTH: 610
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-924037

Query Match 4.5%; Score 45.4; DB 12; Length 610;
Best Local Similarity 53.1%; Pred. No. 29;
Matches 94; Conservative 1; Mismatches 82; Indels 0; Gaps 0;

```
Qy        709 ATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATTTTTACAGTAAAAATATA 768
              ||||||  |  |||  |||    ||    ||  |||||   ||   |||||    ||  |  ||
Db        189 ATTTTATTAAATTATATATAATTCATATAATTACATATGAATTTTACTAGGGATAGGTTA 248

Qy        769 TCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGACCATGGTTATATACAACT 828
              | ||||  ||  | |  ||   ||    |||  || | |   |   ||    | |  |
Db        249 GTATGAAAAACAGTTTATAGAAGGTTTACATTAATTGAAAAATAACCTTTTGTCATACTT 308

Qy        829 TTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATTTATCTGAATTTTAT 885
```

```
                |||  |||  |:  |||       |||  ||  ||||   ||||||  |||||  || |
Db        309 TTGTTTATTTCRCCTTACAAGCTTAGGCATTTAAAACTTGATTTACCTGAAAGTTGT 365
```

RESULT 10
US-09-925-065A-63645/c
; Sequence 63645, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63645
; LENGTH: 721
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-63645

Query Match 4.3%; Score 43.2; DB 7; Length 721;
Best Local Similarity 47.7%; Pred. No. 79;
Matches 123; Conservative 1; Mismatches 134; Indels 0; Gaps 0;

```
Qy        692 AAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATT 751
              || || |    | |   || | ||   |||||||       | || ||    ||   ||
Db        601 AAAGATATTTTATATATATGTAAAGATATTTTATATATATGTAAAGATATTTTATATATA 542

Qy        752 TTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGAC 811
              | || ||  |   ||||| | || || ||   |:|  | || | |   || |  | |||
Db        541 TATAAAGATATTTTATATATATATAAAGATATTWTATATATATATATATATATAAAAGAG 482

Qy        812 CATGGTTATATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATT 871
                |||||||  |  |  ||  |  ||   || || ||  ||  | |||     ||| |
Db        481 TCTGGTTATTGAAGAACTTATCATTGATCTGCTAAGGTCATTTTTATGGTGGGAGATACA 422

Qy        872 TATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATATGATCAACACTGTACTTCGAT 931
              |  | || | |||    | |   | | ||   |   |   | | |   |   | ||
Db        421 CACATAAAATGTATATTTGGTTCCCTGGCAAAGGAGTGAATTTGAGAATACTGTGTTGAA 362

Qy        932 TTGGGAAACTAGCATATG 949
              |||| | |  |   | ||
Db        361 TTGGAAGATCAAACTCTG 344
```

RESULT 11
US-10-301-480-164883/c
; Sequence 164883, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 164883
; LENGTH: 721
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-164883

Query Match 4.3%; Score 43.2; DB 11; Length 721;
Best Local Similarity 47.7%; Pred. No. 79;
Matches 123; Conservative 1; Mismatches 134; Indels 0; Gaps 0;

```
Qy        692 AAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATT 751
              || || |    | |   || | ||   |||||||       | || ||    ||   ||
Db        601 AAAGATATTTTATATATATGTAAAGATATTTTATATATATGTAAAGATATTTTATATATA 542

Qy        752 TTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGAC 811
              | || ||  |   ||||| | || || ||   |:|  | || | |   || |  | |||
Db        541 TATAAAGATATTTTATATATATATAAAGATATTWTATATATATATATATATATAAAAGAG 482

Qy        812 CATGGTTATATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATT 871
                |||||||  |  |  ||  |  ||   || || ||  ||  | |||     ||| |
Db        481 TCTGGTTATTGAAGAACTTATCATTGATCTGCTAAGGTCATTTTTATGGTGGGAGATACA 422

Qy        872 TATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATATGATCAACACTGTACTTCGAT 931
              |  | || | |||    | |   | | ||   |   |   | | |   |   | ||
Db        421 CACATAAAATGTATATTTGGTTCCCTGGCAAAGGAGTGAATTTGAGAATACTGTGTTGAA 362

Qy        932 TTGGGAAACTAGCATATG 949
              |||| | |  |   | ||
Db        361 TTGGAAGATCAAACTCTG 344
```

RESULT 12
US-10-301-480-778292/c
; Sequence 778292, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 778292
; LENGTH: 721
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-778292

Query Match 4.3%; Score 43.2; DB 12; Length 721;
Best Local Similarity 47.7%; Pred. No. 79;
Matches 123; Conservative 1; Mismatches 134; Indels 0; Gaps 0;

```
Qy        692 AAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAATTAGCTAATGATT 751
              || || |    | |   || | ||   |||||||       | || ||    ||   ||
Db        601 AAAGATATTTTATATATATGTAAAGATATTTTATATATATGTAAAGATATTTTATATATA 542

Qy        752 TTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGCTAAAAGAGAGAC 811
              | || ||  |   ||||| | || || ||   |:|  | || | |   || |  | |||
Db        541 TATAAAGATATTTTATATATATATAAAGATATTWTATATATATATATATATATAAAAGAG 482

Qy        812 CATGGTTATATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATGAAAAAAGAGATT 871
                |||||||  |  |  ||  |  ||   || || ||  ||  | |||     ||| |
Db        481 TCTGGTTATTGAAGAACTTATCATTGATCTGCTAAGGTCATTTTTATGGTGGGAGATACA 422

Qy        872 TATCTGAATTTTATGCAAAGCTAGGTTGTAATAAAAATATGATCAACACTGTACTTCGAT 931
```

```
Db        421 CACATAAAATGTATATTTGGTTCCCTGGCAAAGGAGTGAATTTGAGAATACTGTGTTGAA 362

Qy        932 TTGGGAAACTAGCATATG 949
              |||| | |  |   | ||
Db        361 TTGGAAGATCAAACTCTG 344
```

RESULT 13
US-11-079-463-1998
; Sequence 1998, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 1998
; LENGTH: 930
; TYPE: DNA
; ORGANISM: B.fragilis
US-11-079-463-1998

Query Match 4.3%; Score 43; DB 18; Length 930;
Best Local Similarity 49.6%; Pred. No. 90;
Matches 171; Conservative 0; Mismatches 165; Indels 9; Gaps 2;

```
Qy          8 TATCATTTTGTATTCCAACTTATAATAGAAAACAATATCTTGAAGAGTTGTTGAATAGTA 67
              |||| |   |||||   ||||| ||| ||| |||  | |||||     | ||   |
Db         32 TATCTATAGTAATTCCGGTATATAACAGAGAACTATACATAGAAGATGCTATAAAATCTG 91

Qy         68 TAAATAATCAGGAAAAATTTAATTTAGATATTGAGATATGTATATCAGATAATGCCTCTA 127
                | || ||     ||   ||  | ||||||||||   ||||   |||||      ||
Db         92 CTATTAGTC------AAACGTATCAAAATATTGAGATTATTATAGTTGATAACTGTAGTA 145

Qy        128 CTGATGGTACAGAGGAAATGATTGATGTTTGGAGGAACAATTATAATTTCCCAATAATAT 187
              | |||   ||   |   ||  |  ||   ||| | |||  || ||      | |||||
Db        146 CAGATTTAACTTGGAGTATTTTGAATACATGGGCAAGCAAAGATCATAGAATTAAAATAT 205

Qy        188 ATCGGCGTAATAGCGTTAACCTTGGGCCAGATAGGAATTTTCTTGCTTCAGTATCCCTTG 247
              ||| |  ||| |    |||  | || || | |   ||||     ||  |    |  | ||
Db        206 ATCAGAATAAAAC---TAATATAGGACCTGTTTTAAATTGGAATGAATGCTTTAAGCATG 262

Qy        248 CGAATGGGGATTATTGTTGGATATTTGGCAGTGATGATGCTCTTGCGAAAGACTCGTTAG 307
              |    || || |||     |||  || |   |||||||    | ||   ||| |   | |
Db        263 CTTCAGGTGAGTATATAAAGATTCTTTGGTCTGATGATTGGATGGCATTAGATTTTATTG 322

Qy        308 CGATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTAT 352
              | |  |  | |   | |||   |  |||   |   | |||
Db        323 AAAGAGCAGTGAATCTAATTGATGAGAAGTCAGCGTTTGTTATAT 367
```

RESULT 14
US-10-995-561-13443/c
; Sequence 13443, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13443
; LENGTH: 49979
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(49979)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-995-561-13443

Query Match 4.3%; Score 42.8; DB 10; Length 49979;
Best Local Similarity 44.0%; Pred. No. 2.2e+02;
Matches 176; Conservative 1; Mismatches 223; Indels 0; Gaps 0;

```
Qy        500 TTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGGGATGCCATTGATTTTGATG 559
              |||  ||| |||  |   |||||   | | || |   |||  ||      | || |  ||
Db      28878 TTACATATGTAATATAAATAATATAGATATAATATATATGTAATATATATTATATATATT 28819

Qy        560 CGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGCGTATTTAATACGC 619
                |   |||      ||       | || |||||||  ||    |   ||| |||
Db      28818 TATTAATTATATATTATTATATATATTATATATTTATTATATATATTATATATAAATAAT 28759

Qy        620 CAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTAATATGCCGAGGAGATAATGATA 679
                  |  | |  |||||||||| || |  |    ||  ||    |  | |||   |||
Db      28758 ATAAATATAATATATATTATATATAAATAATATAAATATAATATGTATTATATATAAATA 28699

Qy        680 GTTTCGAGAAGAAAGGAAAGGCCAGACGAATTTTAATTGATTTTATTGCATATTTAAAAT 739
              | |  | |  | |   |  |||    || | || |  || | | |       || |
Db      28698 ATATAAATATAATATATATAACCAATATAAATATATTACATATAAATAATATAATATATA 28639

Qy        740 TAGCTAATGATTTTTACAGTAAAAATATATCTTTAAAACGAGCATTTGAAAATGTTTTGC 799
              ||    || | |  || |   | |||||| || |  | |      |  | ||| |
Db      28638 TAATATATAAATAATATAAATATAATATATATTACATATGTAATATAAATAATATAGATA 28579

Qy        800 TAAAAGAGAGACCATGGTTATATACAACTTTGGCTATGGCATGTTATGGCAATAGTGATG 859
              ||| | | |   ||  |||    |  ||   ||        |     |      ||
Db      28578 TAATATATATGTAATATATATTATATATATTTATTAATTATATATTATATATATAAAATA 28519

Qy        860 AAAAAAGAGATTTATCTGAATTTTATGCAAAGCTAGGTTG 899
              | | | | :||| | |  |      | |  |||| ||||
Db      28518 TATATATATWTTTTTTTAGAGACAGCGTATTGCTATGTTG 28479
```

RESULT 15
US-10-301-480-564593/c
; Sequence 564593, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identifiction and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 564593
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-564593

Query Match 4.2%; Score 42.4; DB 12; Length 519;
Best Local Similarity 45.1%; Pred. No. 1e+02;
Matches 157; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

```
Qy        310 ATATTACAAACTTATCTCGATTCTCAAGCAGATATATATTTATGTGACAGAAAAGAGACC 369
              ||||   | |  |   |  ||  |  |  | ||||   | ||| | | | | | |
Db        412 ATATATTATATATTCTTATATAATATATTATATATTCTTATATATAATATATATTATGTA 353
```

```
Qy        370 GGGTGTGATTTAGTTGAGATTAGAAACCCTCATCGTTCTTGGCTCAGAACAGATGATGAA 429
              |   || || |  | ||||   |  || ||   |  |   |    | | || ||  |
Db        352 TTTTTATATATAATATATATTATGTATTCTTATATATTATATATTCTTATATATAATATA 293

Qy        430 CTTTATGTGTTTAATAATAATTTAGATAGGGAAATCTATCTCAGTAGATGCTTATCTATT 489
               | ||| |  |||   ||||| || ||     | || ||    | || ||  |  | |||
Db        292 TTATATATTCTTATATATAATATATATTCTTATATATAATATATTATATATTCTTATATG 233

Qy        490 GGTGGTGTATTTAGCTATCTAAGTTCTTTAATAGTAAAAAAAGAACGATGGGATGCCATT 549
              | | ||    | |   |  |  |  ||  || |     |    |   ||    |
Db        232 TAATATATTTTATATTCTTATATATAATATATTTTATATTCTTATATGTAATATATTTTA 173

Qy        550 GATTTTGATGCGTCCTATATTGGCACTTCCTATCCTCATGTATTTATCATGATGAGCGTA 609
               ||| | ||  ||  ||||||     ||| |||    |     || |    ||     ||
Db        172 TATTCTTATATGTAATATATTTTATATTCTTATATGTAATATATTTTATATATTCTTATA 113

Qy        610 TTTAATACGCCAGGGTGCCTTTTGCATTATATATCAAAACCACTCGTA 657
              | |||||    |      ||| |  || ||||| | |    ||  ||
Db        112 TATAATATATTATATATTCTTATATATAATATATTATATATTCTTATA 65
```

Search completed: May 12, 2006, 06:55:14
Job time : 512.518 secs

THIS PAGE BLANK (USPTO)